

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 02:12:21 ; Search time 5404 Seconds  
(without alignments)  
11427.644 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagaggctgtggtggaag.....tgacatttgggattcccat 2068

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1474.2	71.3	3457	11 AK080397	AK080397 Mus muscu
2	1386.8	67.1	2136	11 AK080680	AK080680 Mus muscu
3	815.4	39.4	1800	14 Y16973	Y16973 RNY16973 Ra
4	711	34.4	896	13 BQ214727	BQ214727 AGENCOURT

5	675.6	32.7	894	14	CD359064	CD359064
6	675.4	32.7	788	12	BG502871	BG502871
7	670.8	32.4	682	14	CB150625	CB150625
8	666	32.2	733	12	BI561523	BI561523
9	653.6	31.6	996	10	BF203684	BF203684
10	646	31.2	757	10	BE734354	BE734354
11	637.4	30.8	1037	12	BG428305	BG428305
12	628.6	30.4	889	13	BU510623	BU510623
13	627.2	30.3	1122	13	BQ066201	BQ066201
14	617.8	29.9	671	9	AV703279	AV703279
15	614.6	29.7	736	12	BI464721	BI464721
16	613	29.6	828	13	BU610641	BU610641
17	588.4	28.5	645	13	BQ638571	BQ638571
18	582.2	28.2	990	12	BG167061	BG167061
19	574	27.8	580	12	BM564978	BM564978
20	571.4	27.6	753	14	CB266771	CB266771
21	566.8	27.4	769	14	CF738084	CF738084
22	566.6	27.4	2659	11	BC037959	BC037959
23	566.4	27.4	568	9	AL708660	AL708660
24	553.4	26.8	828	12	BI761160	BI761160
25	550.4	26.6	742	10	BE870692	BE870692
26	548	26.5	550	9	AL597171	AL597171
27	547.8	26.5	551	12	BM786521	BM786521
28	547.2	26.5	727	12	BI549394	BI549394
29	542.4	26.2	563	10	AW408066	AW408066
30	536.8	26.0	564	10	AW958173	AW958173
31	532.6	25.8	651	14	CF908118	CF908118
32	520.4	25.2	697	12	BG611364	BG611364
33	517	25.0	535	9	AA213789	AA213789
34	503.8	24.4	835	12	BI601238	BI601238
35	500.8	24.2	764	13	EX101439	EX101439
36	500.6	24.2	1065	13	BQ070706	BQ070706
37	498.6	24.1	609	14	CF902795	CF902795
38	497	24.0	505	14	CF138931	CF138931
39	493.6	23.9	712	13	BY733332	BY733332
40	493.4	23.9	1011	13	BU151651	BU151651
41	492	23.8	612	13	BQ771414	BQ771414
42	488.4	23.6	498	10	BF110176	BF110176
43	487.4	23.6	497	10	BF221544	BF221544
44	483	23.4	654	14	CF737683	CF737683
45	482.8	23.3	498	12	BG149493	BG149493

ALIGNMENTS

RESULT 1	AK080397	AK080397	Mus musculus	3 days neonate thymus	cdNA	RIKEN full-length enriched library, clone:A630092F02	product:RRN3 homolog [Homo sapiens], full insert sequence.
LOCUS	AK080397	AK080397	3457 bp	mRNA	linear	HTC 19-SEP-2003	
DEFINITION	AK080397	AK080397	AK080397.1	GI:26099204			
ACCESSION	AK080397	AK080397	HTC; CAP trapper.				
VERSION	AK080397	AK080397	Mus musculus (house mouse)				
KEYWORDS	AK080397	AK080397	Mus musculus				
SOURCE	AK080397	AK080397	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
ORGANISM	AK080397	AK080397	1				
REFERENCE	AK080397	AK080397	1				
AUTHORS	AK080397	AK080397	Carninci,P. and Hayashizaki,Y.				
TITLE	AK080397	AK080397	High-efficiency full-length cDNA cloning				
JOURNAL	AK080397	AK080397	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	AK080397	AK080397	99279253				
PUBMED	AK080397	AK080397	10349636				
REFERENCE	AK080397	AK080397	2				
AUTHORS	AK080397	AK080397	Carninci,P. and Hayashizaki,Y.				
TITLE	AK080397	AK080397	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	AK080397	AK080397	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	AK080397	AK080397	20499374				
PUBMED	AK080397	AK080397	11042159				

REFERENCE  
AUTHORS  
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4  
AUTHORS  
5 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
6  
TITLE  
7 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3457)  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
8  
TITLE  
9 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers  
1. .3457  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:A630092F02"  
/db\_xref="MGI:2407489"  
/db\_xref="taxon:10090"  
/clone="A630092F02"  
/tissue type="thymus"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="3 days neonate"  
14. .1987  
misc\_feature  
/note="PRN3 homolog [Homo sapiens] (SPTR|Q9NYV6, evidence: FASTY, 87.8%ID, 92%length, match=1791)  
putative"

ORIGIN

Query Match 71.3%; Score 1474.2; DB 11; Length 3457;  
Best Local Similarity 85.3%; Pred. No. 0;  
Matches 1704; Conservative 0; Mismatches 283; Indels 10; Gaps 5;  
QY 63 GGCCAATGGCGGCACCGCTGCTTCACACGCGTTCGCGGGAGATGCGCGCGCTTCGTCC 122  
Db 8 GGCCGAATGGCGGCCCGCTGCTTCACACGCGCTTCGCGGGAGATGACGGCAGCGGCC 67  
QY 123 TCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATCGTGCATTAGAG 182  
Db 68 TCTGCC-----ACGCTGAGTGCCTCGCGGACTGGGCTTCCGATATGCTCGCGTTAGAG 121  
QY 183 AATGACTTTTTCAAATCTCCCCCAAGAAAAAACTGTTGCGTTTGGTGGAACCTGTGACAGAA 242  
Db 122 AGCGATTTCTTCAATCTCCTCCGAAAAAACTGTTGCGTTTGGCGGAACCTGTGACAGAA 181  
QY 243 GTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTG 302  
Db 182 GTCTTACTGAAATACAAAAAGGGTGAAACAAATGACTTAGAGTTGTTGAAGAACCCAGCTG 241  
QY 303 TTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAAATTCGGTTCTTCT 362  
Db 242 TCTGATCCTGATATAAAGGATGACCAGATCATTAACCTGGCTACTAGAAATTCGTTCTCT 301  
QY 363 ATCATGTACTTGACAAAAAGACTTTGAGCAACTTATCAGTATTATATAAGATTGCCTTGG 422  
Db 302 GTCATGTACTTGACAAAAAGACTTTGAGCAACTTATAAACATCATATTTGAGATTGCCCTGG 361  
QY 423 TTGAATAGAAGTCAAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTCGTAATCTTGATCA 482  
Db 362 TTGAATAGAAGTCAGAGAGTGGTGAGGAGTATTTGGCTTTTCTTGGTAATCTTGTGCT 421  
QY 483 GCACAGACTGTTTTCCTCAGACCGGTCTCAGCATGATGCTTCCCATTTTGTGCTCC 542  
Db 422 GCACAGACTGTTTTCCTTAGACCATGTCTCAGCATGATGCGTCTCATTTTGTACCTCC 481  
QY 543 CGAGTGATCATTAAGGAAGCGATGTAGATGTTTTCAGATTTCTGATGATGAAGATGATAAT 602  
Db 482 CGAGTAATTTCAAGGAAGGTGGCATAGATGTTTTCAGATTTCTGATGACGAAGATGACAAT 541  
QY 603 CTTCTGCAAAATTTTGACACATGTACAGAGCGCTTGCAAAATAATAGCAAGATATGTACCA 662  
Db 542 CTTCTGCAATTTTGAACATGTACAGAGCGCTTGCAAAATAATAGCAAGATATGTCCCA 601  
QY 663 TCGACACCGTGTTTCTCATGCCAATACTGGTGGAATAATTTCCATTTGTTGGAATACTA 722  
Db 602 TCGACACCATGGTTTCTAATGCCAATACTGGTGGAATAATTTCCGTTTGTGAGGAAGTCC 661  
QY 723 GAGAGAACACTGGAATGTTACGTTTACATACTTACTAAGGATTTAGTGTATATTTTCCAACC 782  
Db 662 GAGAGAACATTTGGAATGTTATGTTTACATACTTATTAAGGATAAGTTTATATTTCCCACT 721  
QY 783 TTGAGGCATGAAATTTCTGGAGCTTATTTGAAAAAATACTCAAGTTTGA -TGTGAATGC 841  
Db 722 TTGAGGCGTGAAATTTCTGGAGCTTGTATTGAAAAAGCTACTCAAGTTAGACTGTGAGTGT 781  
QY 842 ATCCCGGCAGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTTGGGACAGA 901  
Db 782 ATCCCGGCAGGATATTGAAGATGCTGAAGAGAAAGCAGCTCAGACTTGTGTTGGGACAGA 841  
QY 902 TTCCACGGAAGGATGTTTAAATATGGATGAAGATGAAGAAACAGCAACTGAAACAAAGGC 961  
Db 842 TACCACAGAAGGACTGTTTAAATATGGATGAAGATGAGGACACTGACCCCTGAGAAGAAAGC 901  
QY 962 TGGTCTGAAACGGCTCGACCGATGGTGCATCTCTGTAGCCGAGCGCCTGGACATCCTGAT 1021  
Db 902 TGACCAGGAGCAGCCTAACCCAGATGGCTCATCCACCGCAGAGCGCCTGGACGCTCCTGCT 961  
QY 1022 GTCTTTGTTTGTCTTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTTGATAACGG 1081  
Db 962 GTGCTTGCTGCTCTTACATTTGAGGATGTCTGCGGTGTGCACGGTAAAAATTGATAACAA 1021

QY	1082	CAAAACAAAGGATCTATATCGGACCTGATAAAACATCTTTGACAAACTCCTGTGCCAC	1141
Db	1022	TAAACAAAGGATTATACCGTGATCTGATCCATCTTTGACAAACTTCTGTGCCAC	1081
QY	1142	CCATGCCCTCCTGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAATTTGGGATT	1201
Db	1082	ACATGCCCTCCTGCCATGTACAGTTCTTTCATGTTTTTCTCTGCAGCTTCAAGTTGGGATT	1141
QY	1202	CGCAGAGGCATTTTTTGGAAACATCTCTGGAAAAAATTGCAGGACCCAAAGTAATCCTGCCAT	1261
Db	1142	TGCAGAAGCATTTTTTGGAAACATCTCTGGAAAAAAGTTGCAGGATCCAAAATAACCCGCCAT	1201
QY	1262	CATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTTGGCAAGAGCTAAATTTATTCC	1321
Db	1202	CATCAGGCAGGCTGCTGCAAAATTATATTGGGAGCTTTTTTGGCCAGAGCTAAATTTATTCC	1261
QY	1322	TCTTATTACTGTAAATCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAA	1381
Db	1262	TCTTATCACTGTGAACATGCCCTGGATCTCTTGGTTAACTGGCTGCACATGACCTTAC	1321
QY	1382	TAACCAGGATTCGGGAACAAAGGCATCTCGCATGTTGCTCTCCATGGACCATTTTACTC	1441
Db	1322	TAACCAGGATTCGGGAACAAAGGCTTTTGTGACGTTGCACCTCCATGGACCATTTTATTC	1381
QY	1442	AGCCTGCCAAGCTGTCTTCTACACCTTTGTCTTTTAGACACAAGCAGCTTTTGGCGGAAA	1501
Db	1382	AGCCTGCCAAGCTGTCTTCTACACTGTGTCTTTTAGACACAAGCAGCTTTTGGTGAAA	1441
QY	1502	CCTGAAGAAGGTTTGAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCA	1561
Db	1442	CTTGAAGCAAGGTTCTACAGTATCTTCAGAGTCTAAATTTGAGCGCATGTTGTTGAGCA	1501
QY	1562	GCTAATCCCTGAAGATTGGCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAAA	1621
Db	1502	GCTGAACCCACTGAAGATCTGCCTGCCCAAGTGGTTAACTTTCTTGTCTATCACAAA	1561
QY	1622	TAAGTACAGCTCGTCTTCTGTCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCC	1681
Db	1562	TAAATACAGCTGGTGTCTGTCTACACCATCATTTGAGAGGAACAAGTCGCCAGATGCTCCC	1621
QY	1682	AGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACAC	1741
Db	1622	CGTTATCCGAGACACCGCTGGTGAGACTCCGTGCAGACTGCACCAACCCACTGGACAC	1681
QY	1742	CTTCTTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTA	1801
Db	1682	TTTCTTCCCTTTGACCCCTGTGTGCTTAAAGAGGTCAAAGAAATTCATTGATCCTATTTA	1741
QY	1802	TCAGGTATGGGAAGA-CATGAGTGTGAAGAGCTACA-GGAGTTCAAGAAACCCCATGAAA	1859
Db	1742	TCAGATTTGGGAAGATGGGGAGTGTCTGAAGAGCTTCAGGGAGTTTAAAGAAATCTACTAAA	1801
QY	1860	AAGGACATAGTGAAGATGAAGATGATGACTTCTGAAAGCGGAAGTGCCCCAGAAATGAT	1919
Db	1802	AAGGAGGTAGTGGAGGATGAAGATGATGACTTTTGAAGCGGAGGTGCCCCAGAGTGAC	1861
QY	1920	ACCGTGATTGGGATCACACCAAGCTCCTTTGACAGCGATTTCCGAAGTCCCTCAAGTAGT	1979
Db	1862	ACAGTGACTGGCCTTACTCCGAGCTCCTTTGATACCCACTTCCAAAGTCCCTCCAGTAGT	1921
QY	1980	GTGGGCTCCCCACCCGTTGTGTACAT-GCAACCCAGTCCCTCTGACGGGAGAAATTTGT	2038
Db	1922	GTGGGCTCCCCCTCTGTGTGTATATACAGGCCAGTCTCCACTCCTCACAGGATCTAT	1981
QY	2039	GACTGAGATGTGACATT	2055
Db	1982	GATTGAGCTGCGCAATT	1998

RESULT 2  
AK080680  
LOCUS AK080680 2136 bp mRNA linear HTC 19-SEP-2003

DEFINITION	Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830089A07 product:RRN3 homolog [Homo sapiens], full insert sequence.
ACCESSION	AK080680
VERSION	AK080680.1 GI:26099414
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
	2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
	3 Carninci,P., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
	6 (bases 1 to 2136) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/. Location/Qualifiers 1. .2136 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:A830089A07" /db_xref="MGI:2409380" /db_xref="taxon:10090" /clone="A830089A07" /tissue_type="cortex" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days neonate" 40. .1905 /note="RRN3 homolog [Homo sapiens] (SPTR Q9NYV6, evidence: FASTY, 87.8%ID, 92%length, match=1791) putative" 2110. .2115 /note="putative" 2136 /note="putative"									
misc_feature									
polyA_signal									
polyA_site									
ORIGIN									
Query Match 67.1%; Score 1386.8; DB 11; Length 2136; Best Local Similarity 84.5%; Pred. No. 0; Matches 1608; Conservative 0; Mismatches 282; Indels 12; Gaps 4;									
QY	31	TCCGGCCTGAGCGCAGCGGTTCGCGTTAGTTTCGGCCCAATGGCGGACCGCTGCTTCACA	90						
Db	2	TCCGGCGCAAGCGCGGCGTTCGCTCAGTTCAGGCGCGAATGGCGGCCGCTGCTTCACA	61						
QY	91	CGCGTTGCCGGAGATCGCGCGCTTCGTCTCTGCAGTTAAGAAGCTGGCGGCTCGA	150						
Db	62	CGCGCTGTGCGGCGATGTGACGGCAGCGCCTCTGC-----CACGCTGAGTGCCTCGC	115						
QY	151	GGACTGGGATTTCAATATATGCGTGCATTAGAGATGACTTTTTCATTTCTCCCCAAGAA	210						
Db	116	GGACTGGGCTTTCGGATATGCTCGCGTTAGAGCGGATTTCTTCAATTTCTCTCCGAAA	175						
QY	211	AAACTGTTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAGGTGAAA	270						
Db	176	AAACTGTTTCGGTTTGGCGGAACTGTGACAGAAGTCTTACTGAATACAAAAGGTGAAA	235						
QY	271	CAATGACTTTGAGTTGTTGAAGAACCAAGCTGTTAGATCCAGACATAAAGGATGACCAGA	330						
Db	236	CAATGACTTAGAGTTGTTGAAGAACCAAGCTGTCTGATCCTGATATAAAGGATGACCAGA	295						
QY	331	TCATCAACTGGCTGCTAGAATTCGGTTCCTCTCATGTACTTGACAAAAGACTTTGAGC	390						
Db	296	TCATTAACCTGGCTACTAGAATTCGTTCCCTCTGTCATGTACTTGACAAAAGACTTTGAGC	355						
QY	391	AACTTATCAGTATTATATTAAAGATTGCCTTGGTTGAATAGAACTCAAGTCAACAGTAGTGAAG	450						
Db	356	AACTTATAAACATCATATTGAGATTGCCCTGGTTGAATAGAACTCAGAGAGTGGTGGAGG	415						
QY	451	AGTATTTGGCTTTTCTTGGTAACTTTGTATCAGCACAGACTGTTTTCTCAGACCGTGTG	510						
Db	416	AGTATTTGGCTTTTCTTGGTAACTTTGTGTCAGCACAGACTGTCTTCCCTTAGACCATGTC	475						
QY	511	TCAGCATGATTGCTTCCCATTTTGTGCCTCCCGAGTGATCATTAAGGAAGCGATGTAG	570						
Db	476	TCAGCATGATTGCGTCTCATTTTGTACCTCCCGAGTAAATGTCAAGGAAGGTGGCATAG	535						
QY	571	ATGTTTTCAGATTTCTGATGATGAAGATGATAATCTTCTCGTGAATTTTGACATGTACACA	630						
Db	536	ATGTTTTCAGATTTCTGATGACGAAGATGACATCTTCTGCAATTTTGGACATGTACACA	595						
QY	631	GAGCCTTGCAAAATAATAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATAC	690						
Db	596	GAGCCTTGCAAAATAATAACAAGATATGTCCCATCGACACCATGGTTTCTTAATGCCAATAC	655						

QY	691	TGGTGGAAAAAATTTCCATTTGTTTCGAAAAATCAGAGAGAACACTGGAATGTTACGTTTCATA	750						
Db	656	TGGTGGAGAAAATTTCCCGTTTGTGAGGAAGTCCGAGAGAACATTTGGAATGTTATGTTTCATA	715						
QY	751	ACTTACTAAGGATTAGTGTATATTTTCCAACTTTGAGGCATGAAAATTCGGAGCTTATTA	810						
Db	716	ACTTATTAAAGGATAAGTTTATATTTCCCAACTTTGAGCGTGAAAATTCCTGGAGCTTGTTA	775						
QY	811	TTGAAAAAATACTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGTAAGATGCTGAAG	870						
Db	776	TTGAAAAGCTACTCAAGTTAGATGTGAGTGTATCGCGCAGGATATTGAAGATGCTGAAG	835						
QY	871	AAACAGCAAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAATATGGATG	930						
Db	836	AGAAAGCAGCTCAGACTTGTGGTGGGACAGATACCACAGAAGGACTGTTTAATATGGATG	895						
QY	931	AAGATGAAGAAACTGAACATGAACAAAGGCTGGTCTGAAACGGCTCGACACAGATGGTGC	990						
Db	896	AAGATGAGGACACTGACCCCTGAGAAGAAAGCTGACCCAGGAGCAGCCTAAACCAGATGGCTC	955						
QY	991	ATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCTTTGGTTTTTGTCTCTACATGAAGGATG	1050						
Db	956	ATCCCAACCGCAGAGCGCCTGGACGTCTGCTGTGCTGTCTGTCTCTACATTGAGGATG	1015						
QY	1051	TCTGCTATGTAGATGGTAAGTTGATAAATCGGCAAAACAAAGGATCTATATCGCGACCTGA	1110						
Db	1016	TCTGCGGTGTGCACGGTAAAATTGATAACAATAAAACAAAGGATTTATACCGTGATCTGA	1075						
QY	1111	TAAACATCTTTTGACAAAACCTCCTGTTGCCCAACCCATGCCCTCCTGCCATGTACAGTTTTCA	1170						
Db	1076	TATCCATCTTTTGACAAAACCTTCTGTGCCCCACACATGCCCTCCTGCCATGTACAGTTCTTCA	1135						
QY	1171	TGTTTACCTCTGTAGTTTCAAAATGGGATTCGCAGAGGCATTTTGGAAACATCTCTGGA	1230						
Db	1136	TGTTTTCTCTGACAGCTTCAAGTTGGGATTTGCAGAAAGCATTTTGGAAACATCTCTGGA	1195						
QY	1231	AAAAATTGCAGGACCCAAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATTAATATG	1290						
Db	1196	AAAAATTGCAGGATCCAAATAACCCGCCCATCATCAGGCAGGCTGCTGCAAATTAATATG	1255						
QY	1291	GAACTTTTGGCAAGAGCTAAATTTATTCTCTTATTACTGTAAAATCATGCCCTAGATC	1350						
Db	1256	GGAGCTTTTGGCCAGAGCTAAATTTATTTCCTCTTATCCTGTGAAGACATGCCCTGGATC	1315						
QY	1351	TTTTGGTTAACTGGCTGCACATATACCTTAATAACCAAGGATTCGGGAACAAAGCATCT	1410						
Db	1316	TCTTGGTTAACTGGCTGCACATGTACCTTTACTAAACCAAGGATTCGGGAACAAAGCTTTT	1375						
QY	1411	GCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTG	1470						
Db	1376	GTGACGTTGCACT-CATGGACCATTTTATTACGCTGCCAAGCTGTGTTCTACACTGTTG	1434						
QY	1471	TTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAAGTTTGAGTATCTTCAGA	1530						
Db	1435	TTTTTAGACACAAGCAGCTTTTGAGTGGAACCTTGAAGCAAGSTCTACAGTATCTTCAGA	1494						
QY	1531	GTCTGAATTTTGAGCGGATAGTATGAGCCAGCTAAATCCCTGAAAGATTTGCGCTGCCCT	1590						
Db	1495	GTCTAAATTTTGAGCGCAATTTGTTTGAGCCAGCTGAACCCACTGAAGATCTGCGCTGCCG	1554						
QY	1591	CAGTGGTTAACTTTTTTGTGCAATACAAATAAGTACCAGCTCGTCTTCTGCTACACCA	1650						
Db	1555	AAGTGGTTAAATTTCTTTGCTGCTATCACAAATAAATACCAGCTGGTGTCTCTGCTACACCA	1614						
QY	1651	TCATTGAGAGGAACAAATCGCCAGATGCTGCCAGTCAATTAGAG-TACCGCTGGAGGAGAC	1709						
Db	1615	TCATGGAGAGGAACAGTCGCCAGATGCTCCCGTTATCCGCAGCCACCGCTGGTGAGAC	1674						
QY	1710	TCAGTGCAGATCTGCACAAAACCCGCTGGACACCTTCTTCCCTTTGATCCCTGTGTGCTG	1769						
Db	1675	TCCGTGCAGACCTGCACCAACCCACTGGACACTTTCTTCCCTTTTGACCCCTGTGTGCTT	1734						



QY 1770 AAGAGGTCAAAGAAATTCAATTGATCCTATTATCAGGTATGGGAAGACATGAGTCTGAA 1829  
|||||  
Db 1735 AAGAGGTCAAAGAAAGTTTCATTGATCCTATTATCAGATTTGGGAAGATGGGAGTCTGAA 1794  
|||||  
QY 1830 GAGTACAGGAGTTCAAGAAACCCATGAAAAAGGA----CATAGTGAAGATGAAGATGA 1885  
|||||  
Db 1795 GAGCTTCAGGAGTTTAAGAAATCTACTAAAAAGGAATCTCCTGCTGGAAGAAAGGCAATGG 1854  
|||||  
QY 1886 TGACTTTCTGAAAGCGGAAGTGTCCCGAGAATGATACCGTGAT 1927  
|||||  
Db 1855 CACCATTCTGAAGGGAACCTTGTTCTCCAGAAGAGATCTCAAT 1896  
|||||  
  
RESULT 3  
Y16973  
LOCUS Y16973 1800 bp mRNA linear EST 30-APR-1999  
DEFINITION RNY16973 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone  
IRL573, mRNA sequence.  
ACCESSION Y16973  
VERSION Y16973.2 GI:4727035  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 1800)  
AUTHORS Olivier,E., Soury,E., Risler,J.L., Smih,F., Schneider,K.,  
Lochner,K., Jouzeau,J.Y., Fey,G. and Salier,J.P.  
TITLE A novel set of hepatic mRNAs preferentially expressed during an  
acute inflammation in rat represents mostly intracellular proteins  
JOURNAL Genomics 57 (3), 352-364 (1999)  
MEDLINE 99263497  
PUBMED 10329001  
COMMENT On Apr 2, 1998 this sequence version replaced gi:3006069.  
Contact: E. Olivier  
U78 INSERM  
543 chemin de la Breteque, 76233 Bois-Guillaume, France  
POLYA=No.  
  
FEATURES  
source Location/Qualifiers  
1..1800  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="IRL573"  
/tissue\_type="liver"  
/dev\_stage="adult"  
/clone\_lib="Rat liver ESTs (E.Olivier)"  
/note="Organ: liver; Vector: lambda ZAPII; Library  
construction: Ripberger,J.A. et al. J. Biol. Chem. (1995)  
270(50):29998-30006"  
  
ORIGIN  
Query Match 39.4%; Score 815.4; DB 14; Length 1800;  
Best Local Similarity 75.7%; Pred. No. 8.2e-223;  
Matches 1411; Conservative 0; Mismatches 366; Indels 86; Gaps 29;  
  
QY 135 AAGCTGGCGGCTCGAGGACTGGGATTTCAAATATGCGTGCAATAGAGAATCACTTTTC 194  
|||||  
Db 23 ACGCTGAGTGCCTCGGGACTGGGCTTTCAAACATGCTTGCAATAGAGAGTGATTTCTCA 82  
|||||  
QY 195 AATTCTCCCCAAGAAAAAACTGTTTCGGTTTGTGTGGAAGTGTGACAGAAGTCTGCTGAAG 254  
|||||  
Db 83 GTCTCCCAAGAAAC-----TGTCGGTTTGTGTGGACTGTGACAGACGCTCTACTGAGT 133  
|||||  
QY 255 TACAAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACAGACTGTTAGATCCAGAC 314  
|||||  
Db 134 ---ACAAAAGGGTGAACAAATGACTTAGAGTTGTGAAGAA-CAGCTGTCTGATCCTGAT 189  
|||||  
QY 315 ATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCGGTTCTTCTATCATGTACTTG 374  
|||||  
Db 190 ATAGAGGATGATCAGATCATTAACCTGGCTGCTGGAATTCGTTCCCTCTGTCTACTTGTG 249  
|||||

QY 375 ACAAAAGACTTTTGAGCAACTTATCAGTATTATATAAGATTGCCTTGGTTGAATAGAAGT 434  
|||||  
Db 250 ACAAAGACTTTTGAGCAACTTATAAACATCATATGAGATTGCCCTGGTTGAATAGAAGT 309  
|||||  
QY 435 CAAACAGTAGTGGAAAGATATTTGGCTTTTCTTGGTAAATCTTGTATCAGCACAGACTGTT 494  
|||||  
Db 310 CAGACGGTGGTGAGGAGTATTTGGCTTTTCTTGGTAAATCTTGTGTACACAGACTGTG 369  
|||||  
QY 495 TTCCTCAGACCGTGTCTCAGCATGATGCTTCCCATTTTGTGCCCTCCCGAGTGATCAT 554  
|||||  
Db 370 TTCCTTAGACCATGTCTCAGCATGATGCTTCTCATTTTGTACCTCCCGAGTTATTGTG 429  
|||||  
QY 555 AAGGAAGCGGATGTAGATGTTTTCAGATTCTGATGTGAAGATGATAATCTTCCTGCAAA 614  
|||||  
Db 430 AAGGAAGGTGGCATTTGATG-TTCAGAGTCTGAGGATGAAGATGACAAATCTTCCTGCAAA 488  
|||||  
QY 615 TTTGACACATGTACAGAGCCCTTGCAAAATAATAGCAAGATATGTACCATCGACACCGTGG 674  
|||||  
Db 489 TTTGACACATGTACAAAAGCCCTTGCAAAATAATAACAAGATATGTCCCTTCGACACCATGG 548  
|||||  
QY 675 TTTCTCATGCCAATACTGGTGGAAAAAATTTCCATTTGTTTCGAAAAATCAGAGAGAACACTG 734  
|||||  
Db 549 TTTCTAATGCCAATACTGGTAGAAAAAATTTCCATTTTGTGAGGAAATCAGAGAGAACACTG 608  
|||||  
QY 735 GAATGTTACGTTCTATAACTTACTAAGGATTAGTGATATATTTTCCAACTTTGAGGCATGAA 794  
|||||  
Db 609 GAATGCTATGTTCTATAACCTATTAAAGGATAAGTTTATATTTCCCAACGTTGAGGCGTGAA 668  
|||||  
QY 795 ATTCTGAGCTTATTATGTAAAAAACTACTCAAGTTGGATGTGAATGCATCCCGCAGGTT 854  
|||||  
Db 669 ATTCTGAGCTTATTATTGAAAGTC-ACTCAAGTTAGATGTGAGTGTGCGGCGAGGAT 727  
|||||  
QY 855 ATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATTCCACGGAAGGA 914  
|||||  
Db 728 ATTGACGATGCTGAAGAGACAGCAGCTCAGACTTATGTTGGGACAGACGCGCGAGAAGGA 787  
|||||  
QY 915 TTGTTTAAATATGGATGAAGATGAAGAAACTGAACATGAACAAAGGCTGGTCTCTGAACGG 974  
|||||  
Db 788 CTG-TTAAATATGGATGA--GATGAGGAACTGACCCCTGAGAAAGAAAGCTGAGCAGGAGCG 844  
|||||  
QY 975 CTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGCTTTTGGTTTTG 1034  
|||||  
Db 845 CT-GGCCGGATGGCTCATCCTCCACTGCAGAGCGCTGGACATCCTGCTGCTTGTGCTC 903  
|||||  
QY 1035 TCCTACATGAAGGATGTCGTCTATGTAGATGGTAAAGTTGATAACGGCAAAACAAAGGAT 1094  
|||||  
Db 904 TCCTACATTGACGATGTCTGCCATGTGCA--GTAAAAATTGATAACAATAAAACAAAG--A 959  
|||||  
QY 1095 CTATATCGCGACCTGATAAACATCTTTTGACAAAACCTCCTGTTGCCACCCATGCCCTCTGC 1154  
|||||  
Db 960 TTATATCGTGATCTGATATCTATCTT---GACAACTGTGTGTGCCACACATGSCCT-CTGC 1015  
|||||  
QY 1155 CATGTACAGTTTTTTCATGTTTACCTCTGTFAGTTTTCAAATTTGGGATTCGACAGGCAATTT 1214  
|||||  
Db 1016 CATGTACAGTTCTTTCATGTTTTCCTTGCAGC--TCAAATTTGGGTTTGCAGA-GCATT 1072  
|||||  
QY 1215 TTGGAACATCTCTGAAAAAATTTGCAGGACCCCAAGTAATCCTGCCATCATCAGGCAGGCT 1274  
|||||  
Db 1073 TTGGAACATCTCTGAAAAGTGCAGGA----TCAAATAACCCGTCTCATCATCAG--CAGCT 1126  
|||||  
QY 1275 GCTGGAATATATATTGGAAGCTTTTGTGCAAGAGCTAAATTTATTTCCTTTATTACTGTA 1334  
|||||  
Db 1127 GCTGCAAAATACATG---AGCTTTTGTGCAAGAACTAAA--TTAATCCTCTTATCACTGTG 1180  
|||||  
QY 1335 AAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTTCG 1394  
|||||  
Db 1181 AAGTCATGCCTGGATC-TTTGGGTAACTGGCTGCAC-TGTACCTTAATAGCCAGGTTTCGG 1238  
|||||  
QY 1395 GGAACAAAGGCATTTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCT 1454  
|||||  
Db 1239 ACA-----AGGCTTTGTGATGTGCATCATGGACCATTATC---AGCTGCCAAGCT 1287  
|||||

QY	1455	GTGTTCTACACCTTTGTTTTTATGACACACAGCAGCTTTTGTAGCGGAAACCTGAAAGAGGT	1514
Db	1288	GTGT-----CTACACTGTGTTTATGACACAGCACTTT-----GAGTGCACCTGAGGAAGT	1336
QY	1515	TTGACAGTATCTTCAGAGTCTGAAATTTTGTAGCGGATAGTGATGAGCCAGCTAAATCCCTG	1574
Db	1337	CTACGGTATCTCCAGAGTCTAAATTTTGAACGCATTTGTGTCGAGCCAGCTGAACCCACTG	1396
QY	1575	AAGATTTGCTGCGCCTCAGTGGTTAACTTTTTTGTGCTGCAATCAAAATAAGTACCAGCTC	1634
Db	1397	AAGATCTGCCCGCCAGTGTGTATTCCTTGTGTCATCAATAATACAGCTGTGTCGCTA	1456
QY	1635	GTCTTCTGCTACACCATCATTCAGAGGAAACAATCGCCAGATGCTGCCAGTCATTAGGAGT	1694
Db	1457	CACATCATGGAGA-----GGACAATCGCCAGATGCTCCCTGTTAT--CGAGT	1501
QY	1695	ACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAAACCGCTGGACACCTTCTTCCCTTTT	1754
Db	1502	ACAGCTGGTGGAGACTCTGTGCAGACCTGCAACCAACCCATTGGACACTTCTTCCCTTTT	1561
QY	1755	GATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTATGGAA	1814
Db	1562	GACCCCTGTATGCTTAAGAGGTCAAAGAAGTTTCATCGATCCTGTCTATCAGGCTTGGGAA	1621
QY	1815	GACATGAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATAGTGAA	1874
Db	1622	GACATGAGTGATGAAGAGCTGCAGGAGCT--TAGGAACCCACTAGAAAGGAGCTAGTGGAG	1680
QY	1875	GATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGATACCGTGATTGGGATC	1934
Db	1681	GATGAAGACGATGACTTTTGAAGGCGAGTG---CCCAGAGTGACACAGTGGCTGGCCTT	1737
QY	1935	ACACCAAGCTCCTTTTGACACGCATTTCCGAAAGTCTTTCAAGTAGTGTGGGCTCCCCACCC	1994
Db	1738	ACCCCGAACTCCTTTTGACACCCACTTTTCG-AGTCCTTCCAGTAGTGTGGGCTCCCCCTCT	1796
QY	1995	GTG 1997	
Db	1797	GTG 1799	

```

RESULT 4
BQ214727
LOCUS      BQ214727      896 bp      mRNA      linear      EST 02-MAY-2002
DEFINITION AGENCOURT_7595062 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6062849
5', mRNA sequence.
ACCESSION  BQ214727
VERSION    BQ214727.1  GI:20396127
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 896)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1334 row: p column: 18
High quality sequence stop: 595.
Location/Qualifiers
1..896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
source

```

```

/clone="IMAGE:6062849"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match      34.4%; Score 711; DB 13; Length 896;
Best Local Similarity 96.5%; Pred. No. 6.8e-193;
Matches 748; Conservative 0; Mismatches 25; Indels 2; Gaps 2;

QY 1001 CGAGCGCCTGGACATCCTGANGTCTTTGGTTTTTGTCTACATGAAGGATGCTGCTATGT 1060
Db 16 CGCGTCCGTGGACATCCTGATGCTTTGGTTTTTGTCTACATGAAGGATGCTGCTATGT 75

QY 1061 AGATGGTAAGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTT 1120
Db 76 AGATGGTAAGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTT 135

QY 1121 TGACAAACTCCTGTTGCCACCCATGCCCTCCTGCCATGTACAGTTTTTTCATGTTTACCT 1180
Db 136 TGACAAACTCCTGTTGCCACCCATGCCCTCCTGCCATGTACAGTTTTTTCATGTTTACCT 195

QY 1181 CTGTAGTTTCAAATGGGATTTCGAGAGGCCATTTTGGAAACATCTCTGAAAAAATTGCA 1240
Db 196 CTGTAGTTTCAAATGGGATTTCGAGAGGCCATTTTGGAAACATCTCTGAAAAAATTGCA 255

QY 1241 GGACCCAAAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATATATTTGGAAGCTTTT 1300
Db 256 GGACCCAAAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATATATTTGGAAGCTTTT 315

QY 1301 GGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAAATCATGCCATAGATCTTTTGGTTAA 1360
Db 316 GGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAAATCATGCCATAGATCTTTTGGTTAA 375

QY 1361 CTGGCTGCACATATACCTTAATFAACAGGATTTCGGAACAAAGGCATCTCGATGTTGC 1420
Db 376 CTGGCTGCACATATACCTTAATFAACAGGATTTCGGAACAAAGGCATCTCGATGTTGC 435

QY 1421 TCTCCATGAGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTGTGTTTTAGACA 1480
Db 436 TCTCCATGAGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTGTGTTTTAGACA 495

QY 1481 CAAGCAGCTTTTGAGCGGAAACCTGAAAGAGGTTTGACATCTTTCAGAGTCTGAATTT 1540
Db 496 CAAGCAGCTTTTGAGCGGAAACCTGAAAGAGGTTTGACATCTTTCAGAGTCTGAATTT 555

QY 1541 TGAGCGGATAGTGATGAGCGCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAA 1600
Db 556 TGAGCGGATAGTGATGAGCGCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAA 615

QY 1601 CTTTTTTTGTGCAATCAAAATAAGTACCAAGTCCGTCTTCTGCTACA-CCATCAATTGAGA 1659
Db 616 CTTTTTTTGTGCAATCAAAATAAATACCAAGTCCGTCTTCTGCTACACCCATCAATTGAGA 675

QY 1660 GGAAACAATCGCCAGATGCTGCCAGTCAATTAGGAGTACCGCT-GGAGGAGACTCAGTGCAG 1718
Db 676 GGAAACAATCGCCAAATGCTGCCCGTCAATTAAAGAGTACCGCTGGGAGGAGACTCAGTGCAA 735

QY 1719 ATCTGCACAAACCCGCTGGACACCTTCTTCCCTTTTGATCCCTGTGTCTGAAGA 1773
Db 736 ATCTGCACAAACCCGCTGGACACCTTCTTCCCTTTTGGATCCCTGTGTGTGCTGA 790

```

RESULT 5  
CD359064  
LOCUS  
DEFINITION  
ACCESSION

VERSION CD359064.1 GI:31130475  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 894)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM464 row: f column: 24  
High quality sequence stop: 506.  
FEATURES  
source  
1. .894  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30389999"  
/lab\_host="DH10B-Ton A ( T1 and T5 phage resistances)"  
/clone\_lib="NIH MGC 180"  
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;  
Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.68 kb. Library was  
constructed by (Invitrogen). Note: this is a NIH\_MGC  
Library."  
ORIGIN  
Query Match 32.7%; Score 675.6; DB 14; Length 894;  
Best Local Similarity 98.5%; Pred. No. 1.1e-182;  
Matches 703; Conservative 0; Mismatches 9; Indels 2; Gaps 2;  
QY 1152 TGCCATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATTGGGATTCGACAGGCA 1211  
|||||  
Db 1 TGCCATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATTGGGATTCGACAGGCA 60  
QY 1212 TTTTGGACATCTCTGGAAAAAATTGCAGACCCCAAGTAATCTGCCATCATCAGGCAG 1271  
|||||  
Db 61 TTTTGGACATCTCTGGAAAAAATTGCAGACCCCAAGTAATCTGCCATCATCAGGCAG 120  
QY 1272 GCTGCTGGAATTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACT 1331  
|||||  
Db 121 GCTGCTGGAATTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACT 180  
QY 1332 GTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAGGAT 1391  
|||||  
Db 181 GTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAGGAT 240  
QY 1392 TCGGGAACAAGGCATTCTGCGATGTTGCTCTCCATGGACCAATTTTACTACGCTGCCAA 1451  
|||||  
Db 241 TCGGGAACAAGGCATTCTGCGATGTTGCTCTCCATGGACCAATTTTACTCAGCTGCCAA 300  
QY 1452 GCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTTCAGCGGAAACCTGAAAGAA 1511  
|||||  
Db 301 GCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTTCAGCGGAAACCTGAAAGAA 360  
QY 1512 GGTTCAGATATCTCAGAGTCTGAATTTTTCAGCGGATAGTGATGAGCCAGCTAAATCCC 1571  
|||||  
Db 361 GGTTCAGATATCTCAGAGTCTGAATTTTTCAGCGGATAGTGATGAGCCAGCTAAATCCC 420  
QY 1572 CTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATAAGTACCAG 1631

Db 421 CTGAAGATTTGCCTGCCTCAGTGGTTAACTTTTTTGTGCAATCACAAATAAGTACCAG 480  
QY 1632 CTCGTCTTCTGTACACCATCATTGAGAGGAACAATGCCAGATGCTGCCAGTCATTAGG 1691  
|||||  
Db 481 CTCGTCTTCTGTACACCATCTGTGAGAGGAACAATGCCAGATGCTGCCAGTCATTAGG 540  
QY 1692 AGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTCCCC 1751  
|||||  
Db 541 AGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTCCCC 600  
QY 1752 TTTGAT-CCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTATG 1810  
|||||  
Db 601 TTTGATCCCTGTGTGCTGGAGAGTCAAAGAAATTCATTGATCCTATTATCAGGTGTG 660  
QY 1811 GGAAGACATGAGTGTCTGAAGAGCTACAGAGTT-CAAGAAACCCATGAAAAAGG 1863  
|||||  
Db 661 GGAAGACATGAATGCTGAAAAAGCTACAGGAATTCAGAAACCCATGAAAAAAG 714  
RESULT 6  
BG502871 788 bp mRNA linear EST 27-MAR-2001  
LOCUS 602550430F1 NIH\_MGC\_61 Homo sapiens CDNA clone IMAGE:4658071 5',  
DEFINITION mRNA sequence.  
ACCESSION BG502871  
VERSION BG502871.1 GI:13464388  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 788)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1451 row: 1 column: 08  
High quality sequence stop: 699.  
FEATURES  
source  
1. .788  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4658071"  
/tissue\_type="embryonal carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH MGC\_61"  
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:  
SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCCGAGCGCGGACATG-dT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
Library."  
ORIGIN  
Query Match 32.7%; Score 675.4; DB 12; Length 788;  
Best Local Similarity 96.1%; Pred. No. 1.1e-182;  
Matches 756; Conservative 0; Mismatches 23; Indels 8; Gaps 6;

QY	972	CGGCTCGACCATGGTGCATCCTGTAGCGAGCGCCTGGACATCCTGTGTTCTTTGGTT	1031
Db	2	CGGCTCGACCATGGTGCATCCTGTAGCGAGCGCCTGGACATCCTGTGTTCTTTGGTT	61
QY	1032	TTGTCTTACATGAAGGATCTCTGCTATGTAGATGGTAAGTTGATAACGGCAAAACAAAG	1091
Db	62	TTGTCTTACATGAAGGATCTCTGCTATGTAGATGGTAAGTTGATAACGGCAAAACAAAG	121
QY	1092	GATCTATATCGCGACCTGTATAAACATCTTTGACAAACTCCTGTGGCCACCCATGCCTCC	1151
Db	122	GATCTATATCGCGACCTGTATAAACATCTTTGACAAACTCCTGTGGCCACCCATGCCTCC	181
QY	1152	TGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAAATGGGATTCGCAGAGCA	1211
Db	182	TGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAAATGGGATTCGCAGAGCA	241
QY	1212	TTTTTGGAAACATCTCTGGAAAAAATTGCGAGGCCCAAGTAATCCTGCCATCATCAGGCAG	1271
Db	242	TTTTTGGAAACATCTCTGGAAAAAATTGCGAGGCCCAAGTAATCCTGCCATCATCAGGCAG	301
QY	1272	GCTGCTCGAAATTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTTTATTACT	1331
Db	302	GCTGCTCGAAATTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTTTATTACT	361
QY	1332	GTAAATCATGCTAGATCTTTTGGTTAACTGGTGCGATGATACCTTAATAACCCAGGAT	1391
Db	362	GTAAATCATGCTAGATCTTTTGGTTAACTGGTGCGATGATACCTTAATAACCCAGGAT	421
QY	1392	TCGGGAACAAAGGCATCTCGGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGCCAA	1451
Db	422	TCGGGAACAAAGGCATCTCGGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGCCAA	481
QY	1452	GCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAA	1511
Db	482	GCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAA	541
QY	1512	GTTTTCAGTATCTTCAGAGTCTGAAATTTTGAGCGGATAGTGAAGCCAGCTAAATCCC	1571
Db	542	GTTTTCAGTATCTTCAGAGTCTGAAATTTTGAGCGGATAGTGAAGCCAGCTAAATCCC	601
QY	1572	CTGAAGATTTGCCCTGCCTCAGTGGTTAACTTTTTTGTGCAATCACAACAA--TAAGTACCA	1630
Db	602	CTGAAGATTTG-CTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAACAAATTAAGTACCA	660
QY	1631	GCCTGCTCTTCTGCTACACCATCAATT-GAGAGGAACAATCGCCAGA--TGCTGCCAGTCAT	1687
Db	661	GCTCGTCTTCTGTTACACCATCAATTGGAGAGGAACAATCGGAGAATGGTTGCAGTCCAT	720
QY	1688	TAGGAGTACCGCTGGAGAG-ACTCAGTGCAGA--TCTGCACAAACCCGCTGGACACCTT	1744
Db	721	TGGGAGTACCGCTGGAGAGAACTCAGTGCCGAATCTGGGCAAAACCCGGTGAAACCCCT	780
QY	1745	CTTCCCC 1751	
Db	781	TTTTCCC 787	

RESULT 7  
CB150625  
LOCUS  
DEFINITION  
CB150625  
VERSION  
CB150625.1  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 682)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

TITLE  
JOURNAL  
COMMENT

Kim,Y.S.  
21C Frontier Korean EST Project 2001  
Unpublished (2002)  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 32 row: D column: 03  
High quality sequence stop: 682.

FEATURES

source

Location/Qualifiers  
1..682  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="C1SNU17-32-D03"  
/sex="F"  
/tissue\_type="Uterine"  
/cell\_type="Epithelial"  
/cell\_line="SNU-17"  
/lab\_host="Top10F"  
/clone\_lib="C1SNU17"  
/note="Organ: Cervix; Vector: pCNS-D2; Site 1: EcoRI;  
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including  
EcoRI site by treatment of T4 RNA ligase and the first  
strand cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The dt-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F' by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

ORIGIN

Query Match  
Best Local Similarity 32.4%; Score 670.8; DB 14; Length 682;  
Matches 672; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1387 AGAATTCGGGAACAAAGGCATCTGCGATGTTGCTCTCCATGGACCAATTTACTCAGCT 1446  
Db 8 ATGATTCGGGAACAAAGGCATCTGCGATGTTGCTCTCCATGGACCAATTTACTCAGCT 67  
QY 1447 GCCAAGCTGTGTTCTACACCTTTGTTTTTAGACAACAAGCAGCTTTTGAGCGGAAACCTGA 1506  
Db 68 GCCAAGCTGTGTTCTACACCTTTGTTTTTAGACAACAAGCAGCTTTTGAGCGGAAACCTGA 127  
QY 1507 AAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAATGAGCCAGCTAA 1566  
Db 128 AAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAATGAGCCAGCTAA 187  
QY 1567 ATCCCTGAAGATTGCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAATAAAGT 1626  
Db 188 ATCCCTGAAGATTGCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAATAAAGT 247  
QY 1627 ACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCA 1686  
Db 248 ACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCA 307  
QY 1687 TTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCT 1746  
Db 308 TTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCT 367  
QY 1747 TCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGG 1806  
Db 368 TCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGG 427

Qy 1807 TATGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACA 1866  
Db 428 TGTGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACA 487  
Qy 1867 TAGTGGAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCTTCAAGTATGATACCGTGA 1926  
Db 488 TAGTGGAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCTTCAAGTATGATACCGTGA 547  
Qy 1927 TTGGGATCACACCAAGCTCCTTTTGACACCGATTTCCGAAGTCTTCAAGTATGATGATGAGT 1986  
Db 548 TTGGGATCACACCAAGCTCCTTTTGACACCGATTTNCGAAGTCTTCAAGTATGATGATGAGT 607  
Qy 1987 CCCACCCCGTGTGTGTACATGCAACCCAGTCCCTCTGACGGCAGAAAATTTGTGACTGAGA 2046  
Db 608 CCCACCCCGTGTGTGTACATGCAACCCAGTCCCTCTGACGGCAGAAAATTTGTGACTGAGA 667  
Qy 2047 TGTGACATTTGGGAT 2061  
Db 668 TGTGACATTTGGGAT 682

RESULT 8  
BI561523  
LOCUS 603256222F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5298581 5',  
DEFINITION mRNA sequence.  
ACCESSION BI561523  
VERSION BI561523.1 GI:15448837  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 733)  
JOURNAL NIH-MGC <http://mgs.nci.nih.gov/>.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM11755 row: 1 column: 06  
High quality sequence stop: 717.  
Location/Qualifiers  
1. .733

FEATURES  
source  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5298581"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_97"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI  
(gtcgag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTNN-3', size-selected for average  
insert size 2.2 kb and normalized to ROT 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHGRI, National Institutes of Health). Note: this is  
a NIH\_MGC Library."

ORIGIN

Query Match 32.2%; Score 666; DB 12; Length 733;  
Best Local Similarity 97.5%; Pred. No. 5.6e-180;  
Matches 697; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

Qy 53 GCGTTAGTTCGGCCCCAATGGCGGCACCGCTGCTTACACGCGTTCGCCGGGAGATGCGGC 112  
Db 7 GGGTTAGTTCGGCCCCAATGGCGGCACCGCTGCTTACACGCGTTCGCCGGGAGATGCGGC 66  
Qy 113 CGCTTCGCTCCTCTGCAAGTAAAGCTGGGCGCGTCGAGACTGGGATTCAAATATGCG 172  
Db 67 CGCTTCGCTCCTCTGCAAGTAAAGCTGGGCGCGTCGAGACTGGGATTCAAATATGCG 126  
Qy 173 TGCATTAGAGAAATGACTTTTCAATTCCTCCCAAGAAAAAATCTGTCGGTTTGGTGAAC 232  
Db 127 TGCATTAGAGAAATGACTTTTCAATTCCTCCCAAGAAAAAATCTGTCGGTTTGGTGAAC 186  
Qy 233 TGTGACAGAAAGTCTTGTGAAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTTGTGAA 292  
Db 187 TGTGACAGAAAGTCTTGTGAAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTTGTGAA 246  
Qy 293 GAACCCAGCTGTAGATCCAGACATAAAGGATGACCAGATCATCAACTGCTGCTAGAAAT 352  
Db 247 GAACCCAGCTGTAGATCCAGACATAAAGGATGACCAGATCATCAACTGCTGCTAGAAAT 306  
Qy 353 CCGTTCCTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATAAG 412  
Db 307 CCGTTCCTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATAAG 366  
Qy 413 ATTGCCTTGGTTGAATAGAAAGTCAAAACAGTAGTGGAGAGTATTTGGCTTTTCTTGGTAA 472  
Db 367 ATTGCCTTGGTTGAATAGAAAGTCAAAACAGTAGTGGAGAGTATTTGGCTTTTCTTGGTAA 426  
Qy 473 TCTTGATCAGCACAGACTGTTTTCTCAGACCCGTGTCTCAGCATGATTGCTTCCCATTT 532  
Db 427 TCTTGATCAGCACAGACTGTTTTCTCAGACCCGTGTCTCAGCATGATTGCTTCCCATTT 486  
Qy 533 TGTGCCTCCCGAGTGTATTAAGGAAGCGGATGTAGATGTTTCAGATTTCTGATGATGA 592  
Db 487 TGTGCCTCCCGAGTGTATTAAGGAAGCGGATGTAGATGTTTCAGATTTCTGATGATGA 546  
Qy 593 AGATGATAAATCTTCTCGAAAATTTTGACACATGTACAGAGCCCTTGCAATAAATAGCAAG 652  
Db 547 AGATGATAAATCTTCTCGAAA-TTTGACACATGTACAGAGCCCTTGCAATAAATAGCAAG 605  
Qy 653 ATATGTACCATCGACACCGT-GGTTTCTCATGCCAATATCTGGTGGAAAAATTTCCATTTG 711  
Db 606 ATATGTACCATCGACACCGTGGGTTTCTCATGCCAATATCTGGTGGAAAAATTTCCATTT 665  
Qy 712 TTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCAATACTTACTAAGGATTAG 766  
Db 666 GTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCAATACTTACTAAGGATTAG 720

RESULT 9  
BF203684  
LOCUS 601866331F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4099398 5',  
DEFINITION mRNA sequence.  
ACCESSION BF203684  
VERSION BF203684.1 GI:11097270  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 996)  
JOURNAL NIH-MGC <http://mgs.nci.nih.gov/>.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be



found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: L1CM967 row: n column: 07  
High quality sequence stop: 684.  
Location/Qualifiers  
source  
1. .996  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4099398"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC 17"  
/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match		31.6%;	Score 653.6;	DB 10;	Length 996;
Best Local Similarity		97.0%;	Pred. No. 2.5e-176;		
Matches 687;		Conservative 0;	Mismatches 19;	Indels 2;	Gaps 2;
QY	1202	CGCAGAGGCATTTTGGAAACATCTCTGGAAAAAATTTCGAGACCCCAAGTAATCCTGCCAT	1261		
Db	2	CGCAGAGGCATTTTGGAAACATCTCTGGAAAAA--TGCAAGACCCCAAGTAATCCTGCCAT	60		
QY	1262	CATCAGCAGGCTGCTCGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTATTC	1321		
Db	61	CATCAGCAGGCTGCTCGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTATTC	120		
QY	1322	TCTTATTACTGTAAATFCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAA	1381		
Db	121	TCTTATTACTGTAAATFCATGCCCTAGATCTTTGTTGTTAACTGGCTGCACATATACCTTAA	180		
QY	1382	TAACCAAGGATTCGGGAACAAAGGCATTCGCGATGTTGCTCTCCATGGACCATTTTACTC	1441		
Db	181	TAACCAAGGATTCGGGAACAAAGGCATTCGCGATGTTGCTCTCCATGGACCATTTTACTC	240		
QY	1442	AGCCTGCAAGCTGTGTTCTACACCTTTGTTTATAGACACAAGCAGCTTTTGAGCGGAAA	1501		
Db	241	AGCCTGCAAGCTGTGTTCTACACCTTTGTTTATAGACACAAGCAGCTTTTGAGCGGAAA	300		
QY	1502	CCTGAAGAAGGTTTGCAGTATCTTTCAGAGTCTGAATTTTGAGCGGATAGTGATGACCA	1561		
Db	301	CCTGAAGAAGGTTTGCAGTATCTTTCAGAGTCTGAATTTTGAGCGGATAGTGATGACCA	360		
QY	1562	GCTAAATCCCCTGAAGATTTCCTGCCCTCAGTGGTTAACTTTTTTCTGCAATCAGAA	1621		
Db	361	GCTAAATCCCCTGAAGATTTCCTGCCCTCAGTGGTTAACTTTTTTCTGCAATCAGAA	420		
QY	1622	TAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCC	1681		
Db	421	TAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCC	480		
QY	1682	AGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACAC	1741		
Db	481	AGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACAC	540		
QY	1742	CTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTA	1801		
Db	541	CTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTA	600		
QY	1802	TCAGTATGGGAAGACATGAGTCTGTAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAA	1861		
Db	601	TCAGTATGGGAAGACATGAGTCTGTAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAA	660		
QY	1862	GGACATAGTGGGAAGATGAAGATGATGATCTTCTGAAAGCGGAAGTGCC	1909		
Db	661	GGACATAGTGGGAAGATGAAGATGATG-CATTCTGAAAGGAGAAATGGCC	707		

RESULT 10  
BE734354

LOCUS  
DEFINITION  
60156596F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3840623 5',  
mRNA sequence.  
ACCESSION  
BE734354  
VERSION  
BE734354.1 GI:10148346  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 757)  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC

REFERENCE

AUTHORS  
TITLE  
JOURNAL  
COMMENT  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: L1CM530 row: o column: 24  
High quality sequence stop: 693.  
Location/Qualifiers

FEATURES

source

1. .757  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3840623"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_21"  
/note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match		31.2%;	Score 646;	DB 10;	Length 757;
Best Local Similarity		95.9%;	Pred. No. 3.3e-174;		
Matches 727;		Conservative 0;	Mismatches 25;	Indels 6;	Gaps 6;
QY	607	CTGCAAAATTTTGACACATGTCACAGAGCCTTGCAATAATAGCAAGATATGTACCATCGA	666		
Db	1	CTGCAAAATTTTGACACATGTCACAGAGCCTTGCAACTAATAGCAAGATATGTACCATCGA	60		
QY	667	CACCGTG-GTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTTTCGAAAAATCAGAG	725		
Db	61	CACCGTGATTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTTTCGAAAAATCAGAG	120		
QY	726	AGAACACTGGA-ATGTTACGTTTCATAACTTACTAAGGATTAGTGATATATTTTCCACCTT	784		
Db	121	AGAACACTGGACATGTTACGTTTCATAACTTACTAAGGATTAGTGATATATTTTCCACCTT	180		
QY	785	GAGGCATGAAATCTGGAGCTTATATTGAAAAACTACTCAAGTTGGATGTGAATGCATC	844		
Db	181	GAGGCATGAAATCTGGAGCTTATATTGAAAAACTACTCAAGTTGGATGTGAATGCATC	240		
QY	845	CCGCAGGGG-TATTGAAGATGCTGAAGAAACAGCAACTCAAACCTTGTGGTGGACAGATT	903		
Db	241	CCGCAGGGGCTATTGAAGATGCTGAAGAAACAGCAACTCAAACCTTGTGGTGGACAGATT	300		
QY	904	CCACGGAAGGATTGTTTAAATATGGATGAAGATGAAGAAACTGAACATGAACAAA-GGCT	962		

Db 301 CCACGGAAGGATTGTTTAATATGGATGAAGAACTGAACATGAACAAACGCGCT 360

QY 963 GGTCTGAACGGCTCGACCAGATGGTGCATCCTGTPAGCCGAGCGCCTGGACATCCTGATG 1022

Db 361 GGTCTGAACGGCTCGACCAGATGGTGCATCCTGTPAGCCGAGCGCCTGGACATCCTGATG 420

QY 1023 TCTTTGGTTTTTGCTCTACATGAAGGATGTCTGCTATGTAGATGGTAAGTTGATAACGGC 1082

Db 421 TCTTTGGTTTTTGCTCTACATGAAGGATGTCTGCTATGTAGATGGTAAGTTGATAACGGC 480

QY 1083 AAAACAAAGGATCTATATCGCGACCTGTATAAACATCTTTGACAAAACCTCCTGTTGCCACC 1142

Db 481 AAAACAAAGGATCTATATCGCGACCTGTATAAACATCTTTGACAAAACCTCCTGTAGCCCAAC 540

QY 1143 CATGCTCCTGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAATTTGGGATTC 1202

Db 541 CATGCTCCTGCCATGTACAGTATTACATGTATTACCTCTGTAGTTTCAAATCGGGATTC 600

QY 1203 GCAGAGGCATTTTTTGGAACATCTCTGGAATAAAATTGCAGGACCCCAAGTAATCCTGCCATC 1262

Db 601 GCAGAGGCATATTAGGAACATCTCTGGAATAAAATTGCAGGACCCCAAGTAATCCTGCCATC 660

QY 1263 ATCAGGAGGCTGCTGGAATAATTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCT 1322

Db 661 ATCAGGAGGCTGCTGGAATAATTATATTGGAAGC-TTATGGGAAGCGCTAATTAATTCCT 719

QY 1323 CTTATTACTGTAAATCATGCCCTAGATCTTTTGGTTAA 1360

Db 720 C-TATTACTGGTCCATCATGCCTAGAAATCTTAGGATAA 756

RESULT 11

LOCUS BG428305

DEFINITION 602498983F1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4612605 5', mRNA linear EST 14-MAR-2001

ACCESSION BG428305

VERSION BG428305.1 GI:13334811

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 1037)

TITLE NIH-MGC <http://mgc.nci.nih.gov/>.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLCW1361 row: e column: 22

High quality sequence stop: 637.

Location/Qualifiers

FEATURES

source

1..1037

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4612605"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NIH\_MGC\_75"

/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggccgctcgcc); Site\_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts

ORIGIN

Query Match 30.8%; Score 637.4; DB 12; Length 1037;  
Best Local Similarity 88.7%; Pred. No. 1.2e-171;  
Matches 797; Conservative 0; Mismatches 81; Indels 21; Gaps 9;

QY 49 GGTCCGCTTAGTTCGGCCCAATGGCGGCACCGCTGTTTCACACGCGTTTGCCTGGGAGATG 108

Db 2 GGTCCGCTTAGTTCGGCCCAATGGCGGCACCGCTGTTTCACACGCGTTTGCCTGGGAGATG 61

QY 109 CGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCTCGAGGACTGGGATTTCAAATA 168

Db 62 CGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCTCGAGGACTGGGATTTCAAATA 121

QY 169 TCGCTGCATTAGAGAATGACTTTTTCAATTCTCCCAAGAAAACTGTTTCGGTTTGGTG 228

Db 122 TCGCTGCATTAGAGAATGACTTTTTCAATTCTCCCAAGAAAACTGTTTCGGTTTGGTG 181

QY 229 GAACTGTGACAGAAGCTTGTCTGAAGTACAAAAGGGTGAACAAATGACTTTTGAGTTGT 288

Db 182 GAACTGTGACAGAAGCTTGTCTGAAGTACAAAAGGGTGAACAAATGACTTTTGAGTTGT 241

QY 289 TGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAG 348

Db 242 TGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAG 301

QY 349 AATTCGCTTCTTCTATCATGTACTTTGACAAAAGACTTTTGAGCAACTTATCAGTATTATAT 408

Db 302 AATTCGCTTCTTCTATCATGTACTTTGACAAAAGACTTTTGAGCAACTTATCAGTATTATAT 361

QY 409 TAAGATTGCCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCTTG 468

Db 362 TAAGATTGCCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCTTG 421

QY 469 GTAATCTTGATCAGCACAGACTGTTTTCCTCAGACCCGTGCTCAGCATGATTCCTCCC 528

Db 422 GTAATCTTGATCAGCACAGACTGTTTTCCTCAGACCCGTGCTCAGCATGATTCCTCCC 481

QY 529 ATTTTGTGCTCCCG-AGTGATCATTAAGGAAGGCGGATGTAGATGTTTCAGATTCCTGAT 587

Db 482 ATTTTGTGCTCCCGGAAAGTGATCATTAAGGAAGGCGGATGTAGATGTTTCAGATTCCTGAT 541

QY 588 GATGAAGATGATAATCTTCTCCTGCAAAATTTGACACATGTACAGAGCCCTGCAAAATAATA 647

Db 542 GATGAAGATGATAATCTTCTCCTGCAAA-TTTGACACATGTACAGAGCCCTGCAAAATAATA 600

QY 648 GCAAGATATGTACCATCGACACCG-TGGTTTCTCATGCAATACCTGG-TGGAAAAATTT 704

Db 601 GCAAGATATGTACCATCGACACCGTTGTTTCTCATGCAATACCTGGTGGCAACAATTC 660

QY 705 CCATTTGTTTGAATAA---TCAGAGAGAAACACTGGAAATGTT--ACGTTTCAAT--AACTTAC 756

Db 661 CCATATGTTTGAATAAATACTCCAGAGAGAAACACTGGAACTGTCAACCGTTTCAATGAACTTACT 720

QY 757 TAAGGATTAAGTGTATATTTTCCAACTTG--AGGCAATGAAATTTCTGGAGCTTATTATGA 814

Db 721 TCAGGATCAGGGATATGTCCCAAGCTTGGAGGCAATGCAAAATTTCTGGAGCTTATTATGA 780

QY 815 AAAACTACTCAAGTTGGATGTGAATGCATCCCGGACAGGTTTGAAGATG-----CTGA 868

Db 781 AAACTACTCAAGTTGGATGTGAATGCTCCCGGGCGGGGTATTTGAAGACTGCTGAGG 840

QY 869 AGAAACAGCAACTCAAACTTGTGGTGGACAGATTCACCGGAAGGATTTTAATATCG 927

Db 841 AAAGCAGGAAGTACAACTGGTGGCGGAACAATCCCGGCAAGGATTTTTCAGTTTGG 899

RESULT 12

BU510623

LOCUS

889 bp mRNA linear EST 12-SEP-2002



ORIGIN

of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

Query Match 30.3%; Score 627.2; DB 13; Length 1122;  
Best Local Similarity 98.5%; Pred. No. 1.1e-168;  
Matches 644; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

QY 13 GCTGGAAGGAGCTGGGCATCCGCCTGAGGCGCAGCGGTGCGTTAGTTTCGGGCCAATGG 72  
Db 1 GCTGGAAGGAGCTGGGCATCCGCCTGAGGCGCAGCGGTGCGTTAGTTTCGGGCCAATGG 60

QY 73 CGGCACCGCTGCTTCACACGCGCTTCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTTA 132  
Db 61 CGGCACCGCTGCTTCACACGCGCTTCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTCA 120

QY 133 AGAAGCTGGGCGGCTCGAGGACTGGGATTTCAATATGCGTGCATTAGAGAATGACTTTT 192  
Db 121 AGATGCTGGGCGGCTCGAGGACTGGGATTTCAATATGCGTGCATTAGAGAATGACTTTT 180

QY 193 TCAATTTCTCCCAAGAAACCTGTTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGA 252  
Db 181 TCAATTTCTCCCAAGAAACCTGTTTCAGTTTGGTGGAACTGTGACAGAAGTCTTGCTGA 240

QY 253 AGTACAAAAGGGTGAAACAAATGACTTTTGAGTTGTTGAAGAACCGACTGTTAGATCCAG 312  
Db 241 AGTACAAAACGGTGAAACAAATGACTTTTGAGTTGTTGAAGAACCGACTGTTAGATCCAG 300

QY 313 ACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAAATTCGGTTCTTCTATCATGTACT 372  
Db 301 ACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAAATTCGGTTCTTCTATCATGTACT 360

QY 373 TGACAAAAGACTTTGAGCAACTTATCAGTATATATTAAGATTGCCTTGGTTGAATAGAA 432  
Db 361 TGACAAAAGACTTTGAGCAACTTATCAGTATATATTAAGATTGCCTTGGTTGAATAGAA 420

QY 433 GTCAACAGTAGTGAAGAGATTTTGGCTTTTCTTGGTAATCTTGATATCAGCACAGACTG 492  
Db 421 GTCAACAGTAGTGAAGAGATTTTGGCTTTTCTTGGTAATCTTGATATCAGCACAGACTG 480

QY 493 TTTTCTCTCAGACCGTGCTCAGCATGATTGTTGCCATTTTGTGCCCTCCCGAGTGATCA 552  
Db 481 TTTTCTCTCAGACCGTGCTCAGCATGATTGTTGCCATTTTGTGCCCTCCCGAGTGATCA 540

QY 553 TTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTCCTGCAA 612  
Db 541 TTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTCCTGCAA 600

QY 613 ATTTTGACACATGTCACAGAGCCTTGCAATAATAGCAAGATATGTACCATCGA 666  
Db 601 ATTTTGACAC--GTCACAGAGCCTTGCAATAATAGCAAGATATGTACCATTGA 652

RESULT 14  
AV703279  
LOCUS AV703279 671 bp mRNA linear EST 09-OCT-2000  
DEFINITION AV703279 ADB Homo sapiens cDNA clone ADBCCD12 5', mRNA sequence.  
ACCESSION AV703279  
VERSION AV703279.1 GI:10720608  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 671)  
AUTHORS Peng, Y., Song, H., Huang, Q., Gu, Y., Yang, Y., Gao, G.,  
Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z.,  
Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S.,  
Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z. and Han, Z.  
Homo sapiens cDNA ADB clones  
Unpublished (2000)

COMMENT

Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES

Location/Qualifiers  
1..671  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="ADBCCD12"  
/tissue\_type="Adrenal gland"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/clone\_lib="ADB"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

ORIGIN

Query Match 29.9%; Score 617.8; DB 9; Length 671;  
Best Local Similarity 97.0%; Pred. No. 4.1e-166;  
Matches 650; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

QY 416 GCCTTGGTTGAATAGAAAGTCAAAACAGTAGTGGAAAGAGTATTTGGCTTTTCTGGTAATCT 475  
Db 1 GCCTTGGTTGAATAGAAAGTCAAAACAGTAGTGGAAAGAGTATTTGGCTTTTCTGGTAATCT 60

QY 476 TGTATCAGCACAGACTGTTTTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGT 535  
Db 61 TGTATCAGCACAGACTGTTTTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGT 120

QY 536 GCCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGA 595  
Db 121 GCCTCCCGAGTGATCATTAAGGAAGGCGATGTACATGTTTCAGATTCTGATGATGAAGA 180

QY 596 TGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATATAGCAAGATA 655  
Db 181 TGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATATAGCAAGATA 240

QY 656 TGTACCATCGACACCGTGGTTTCTCATGCCAATACCTGGTGGAAAAATTTCCATTTGTTG 715  
Db 241 TGTACCATCGACACCGTGGTTTCTCATGCCAATACCTGGTGGAAAAATTTCCATTTGTTG 300

QY 716 AAAATCAGAGAGAACAACACTGGAATGTTACGTTTCAATACTTACTAAGGATTAGTGATATTT 775  
Db 301 AAAATCAGAGAGAACAACACTGGAATGTTACGTTTCAATACTTACTAAGGATTAGTGATATTT 360

QY 776 TCCAACCTTGAGGCATGAAATCTTGAGCTTATTATTGAAAAACTACTCAAGTTGGATGT 835  
Db 361 TCCAACCTTGAGGCATGAAATCTTGAGCTTATTATTGAAAAACTACTCAAGTTGGATGT 420

QY 836 GAATGCATCCCGGCAGGTTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTTGG 895  
Db 421 GAATGCATCCCGGCAGGTTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTTGG 480

QY 896 GACAGATTCCACGGAAGGATGTTTAAATATGATGAAGATGAAGAACTGAACATGAAC 955  
Db 481 GACAGATTCCACGGAAGGATCGTTTAAATATGATGAAGATGACGAAACTGAACACTCAAC 540

QY 956 AAAGGCTGG-TCCTGAACGGCTCGACCATGTTGGTGCATCTGTAGCCGAGCGCTGGACA 1014  
Db 541 ACAGGCTGNTTCCCTGAACGGCTCGACCATGTTGGTGCATCTGTAGTCGACCGCTGGACA 600

QY 1015 TCCTGATGTTTGGTTTGTCTCCATACATGAAGGATGCTGCTATGTAGATGGTAAGGTTG 1074  
Db 601 TCCTGATGTTTGGTTTGTCTCCATACATGAAGGATGCTGCTATGTAGATGGTAAGGTTG 659

QY 1075 ATAACGGCAA 1084  
Db 660 ATACCGCCCA 669





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 04:52:59 ; Search time 931 Seconds  
(without alignments)  
10836.727 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagaggctgtggctggaag.....tgacattgggattcccat 2068

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2068	100.0	2068	9	US-09-932-678-1
2	2018.4	97.6	2040	9	US-09-952-013A-1
3	1457.6	70.5	1770	16	US-10-094-466-7
4	567	27.4	2493	16	US-10-094-749-544
5	420.8	20.3	1461	13	US-10-221-625-172
6	230	11.1	2103	9	US-09-925-301-192
7	192	9.3	535	15	US-10-029-386-10613
8	188.8	9.1	535	15	US-10-029-386-12002
9	184	8.9	184	15	US-10-029-386-24313
10	182.4	8.8	184	15	US-10-029-386-25702
11	156	7.5	557	9	US-09-998-598-1177
12	153.2	7.4	598	15	US-10-029-386-7910
13	146.6	7.1	225	9	US-09-783-590-11086
14	140.6	6.8	506	15	US-10-029-386-8224

15	139.6	6.8	342	15	US-10-029-386-21924	Sequence 21924, A
16	116.8	5.6	140	9	US-09-728-445-420	Sequence 420, App
17	106.6	5.2	414	9	US-09-864-761-4848	Sequence 4848, Ap
18	93.4	4.5	95	9	US-09-864-761-21583	Sequence 21583, A
19	84.2	4.1	89	15	US-10-029-386-21610	Sequence 21610, A
20	61.6	3.0	634	13	US-10-027-632-263885	Sequence 263885,
21	61.6	3.0	634	13	US-10-027-632-263886	Sequence 263886,
22	61.6	3.0	634	13	US-10-027-632-263887	Sequence 263887,
23	61.6	3.0	634	13	US-10-027-632-263888	Sequence 263888,
24	61.6	3.0	634	16	US-10-027-632-263885	Sequence 263885,
25	61.6	3.0	634	16	US-10-027-632-263886	Sequence 263886,
26	61.6	3.0	634	16	US-10-027-632-263887	Sequence 263887,
27	61.6	3.0	634	16	US-10-027-632-263888	Sequence 263888,
28	61.2	3.0	634	13	US-10-027-632-263889	Sequence 263889,
29	61.2	3.0	634	16	US-10-027-632-263889	Sequence 263889,
30	49.4	2.4	51	13	US-10-393-815-96	Sequence 96, Appl
31	49.4	2.4	51	13	US-10-393-815-97	Sequence 97, Appl
32	48.2	2.3	12770	16	US-10-397-569-4	Sequence 4, Appli
33	47.6	2.3	1671	15	US-10-032-585-6770	Sequence 6770, Ap
34	39.2	1.9	593	13	US-10-424-599-80209	Sequence 80209, A
35	39.2	1.9	1819	13	US-10-424-599-135210	Sequence 135210,
36	39.2	1.9	3108	15	US-10-193-764-68	Sequence 68, Appl
37	39.2	1.9	4937	14	US-10-092-880-3	Sequence 3, Appli
38	39.2	1.9	4937	15	US-10-193-764-66	Sequence 66, Appl
39	39.2	1.9	684707	16	US-10-398-221-9	Sequence 9, Appli
40	39.2	1.9	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
41	38.6	1.9	638	9	US-09-974-300-6328	Sequence 90356, A
42	38.6	1.9	3284	13	US-10-424-599-90356	Sequence 153, App
43	37.8	1.8	2418	15	US-10-349-680-153	Sequence 31, Appl
44	37.8	1.8	2910	15	US-10-193-764-31	Sequence 29, Appl
45	37.8	1.8	2928	15	US-10-193-764-29	

ALIGNMENTS

RESULT 1  
US-09-932-678-1  
; Sequence 1, Application US/09932678  
; Patent No. US20020090706A1  
; GENERAL INFORMATION:  
; APPLICANT: Reeder, Ronald H.  
; APPLICANT: Moorefield, Beth  
; APPLICANT: Greene, Elizabeth A.  
; TITLE OF INVENTION: HUMAN RRN3 AND COMPOSITIONS AND METHODS RELATING  
; TITLE OF INVENTION: THERETO  
; FILE REFERENCE: 14538A-005810US  
; CURRENT APPLICATION NUMBER: US/09/932.678  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: 60/225,893  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2068  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-932-678-1

Query Match	100.0%;	Score 2068;	DB 9;	Length 2068;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2068;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTGAGCGCAGCGGTTCGCTTAGT	60	
Db	1	ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTGAGCGCAGCGGTTCGCTTAGT	60	
QY	61	TCGGCCCAATGGCGCACCGCTGCTTACACGCGTTTCCGGGAGATCGGCCGCTTCGT	120	
Db	61	TCGGCCCAATGGCGCACCGCTGCTTACACGCGTTTCCGGGAGATCGGCCGCTTCGT	120	
QY	121	CCTCTGCAGTTAAGAAGCTGGCGGCTCGAGGACTGGGATTTCAATATCGTGCATTAG	180	

Db 121 CCTCTGCAGTTAAGAACTGGCGCTCGAGGACTGGGATTTCAAATATCGGTGCATTAG 180

QY 181 AGAATGACTTTTTCAATCTCCCCAAGAAAAAAGTCTCGGTTTGGTGGAACTGTGACAG 240

Db 181 AGAATGACTTTTTCAATCTCCCCAAGAAAAAAGTCTCGGTTTGGTGGAACTGTGACAG 240

QY 241 AAGTCTTGCTGAAGTACAAAAAGGGTGAAAAAAGTGAACCTTTGAGTTGTTGAAGAACAGC 300

Db 241 AAGTCTTGCTGAAGTACAAAAAGGGTGAAAAAAGTGAACCTTTGAGTTGTTGAAGAACAGC 300

QY 301 TGTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGTAGAAATTCGGTTCTT 360

Db 301 TGTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGTAGAAATTCGGTTCTT 360

QY 361 CTATCATGTACTTGACAAAAAGACTTTGAGCAACTTATCAGTATTATATAAGATTGCCTT 420

Db 361 CTATCATGTACTTGACAAAAAGACTTTGAGCAACTTATCAGTATTATATAAGATTGCCTT 420

QY 421 GGTTGAATAGAAAGTCAAAACAGTAGTGGAAAGAGTATTTGGCTTTTCTTGGTAATCTTGTAT 480

Db 421 GGTTGAATAGAAAGTCAAAACAGTAGTGGAAAGAGTATTTGGCTTTTCTTGGTAATCTTGTAT 480

QY 481 CAGCACAGACTGTTTTCTCAGACCCGTGTCTCAGCATGATGCTTCCCATTTTGTGCCTC 540

Db 481 CAGCACAGACTGTTTTCTCAGACCCGTGTCTCAGCATGATGCTTCCCATTTTGTGCCTC 540

QY 541 CCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTTCAGATTCCTGATGAAGATGATA 600

Db 541 CCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTTCAGATTCCTGATGAAGATGATA 600

QY 601 ATCTTCTCTGCAAAATTTTGACACATGTTCAGAGCCCTTGCAAAATAATAGCAAGATATGTAC 660

Db 601 ATCTTCTCTGCAAAATTTTGACACATGTTCAGAGCCCTTGCAAAATAATAGCAAGATATGTAC 660

QY 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTTCGAAAAAT 720

Db 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTTCGAAAAAT 720

QY 721 CAGAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGATTAGTGATATTTTCCAA 780

Db 721 CAGAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGATTAGTGATATTTTCCAA 780

QY 781 CCTTGAGGCATGAAATCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTGAATG 840

Db 781 CCTTGAGGCATGAAATCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTGAATG 840

QY 841 CATCCCGCAGGGTATTGAAGATGCTGAAGAAAAACAGCAACTCAAACITTTGGTGGGACAG 900

Db 841 CATCCCGCAGGGTATTGAAGATGCTGAAGAAAAACAGCAACTCAAACITTTGGTGGGACAG 900

QY 901 ATTCCACGGAAGGATTGTTTAATATATGGATGAAGATGAAGAAACTGAACATGAACAAAGG 960

Db 901 ATTCCACGGAAGGATTGTTTAATATATGGATGAAGATGAAGAAACTGAACATGAACAAAGG 960

QY 961 CTGGTCTGAAACGGCTCGACCCAGATGGTGCATCCTGTAGCGGAGCGCTGGACATCCTGA 1020

Db 961 CTGGTCTGAAACGGCTCGACCCAGATGGTGCATCCTGTAGCGGAGCGCTGGACATCCTGA 1020

QY 1021 TGTCTTTGGTTTGTCTACATGAAGGATGCTGCTATGTAGATGGTAAGTTGATAACG 1080

Db 1021 TGTCTTTGGTTTGTCTACATGAAGGATGCTGCTATGTAGATGGTAAGTTGATAACG 1080

QY 1081 GCAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTTTGACAAAACCTCTGTTGCCCA 1140

Db 1081 GCAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTTTGACAAAACCTCTGTTGCCCA 1140

QY 1141 CCCATGCCCTCCTGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAATGGGAT 1200

Db 1141 CCCATGCCCTCCTGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAATGGGAT 1200

QY 1201 TCGCAGAGGCATTTTGGAAACATCTCTCGAAAAAATTTGCAGACCCCAAGTAATCCTGCCA 1260

Db 1201 TCGCAGAGGCATTTTGGAAACATCTCTCGAAAAAATTTGCAGACCCCAAGTAATCCTGCCA 1260

QY 1261 TCATCAGGCAGGCTGCTGGAAAAATATATTTGGAAGCTTTTTTGGCAAGAGCTAAATTTATTC 1320

Db 1261 TCATCAGGCAGGCTGCTGGAAAAATATATTTGGAAGCTTTTTTGGCAAGAGCTAAATTTATTC 1320

QY 1321 CTCTTATTACTGTAAAAATCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380

Db 1321 CTCTTATTACTGTAAAAATCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380

QY 1381 ATAAACCAGGATTCGGGAACAAAGGCATTCCTCGCATGTTGCTCTCCATGGACCATTTTACT 1440

Db 1381 ATAAACCAGGATTCGGGAACAAAGGCATTCCTCGCATGTTGCTCTCCATGGACCATTTTACT 1440

QY 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTCTTGTAGACACAAAGCAGCTTTTGAGCGGAA 1500

Db 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTCTTGTAGACACAAAGCAGCTTTTGAGCGGAA 1500

QY 1501 ACCTGAAAGAAAGTTTGCAGTATCTTCAGAGTCTGAATTTTGTAGCGGATAGTGATGAGCC 1560

Db 1501 ACCTGAAAGAAAGTTTGCAGTATCTTCAGAGTCTGAATTTTGTAGCGGATAGTGATGAGCC 1560

QY 1561 AGCTAAATCCCTGAAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTTTGTCTGCAATCAAA 1620

Db 1561 AGCTAAATCCCTGAAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTTTGTCTGCAATCAAA 1620

QY 1621 ATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGC 1680

Db 1621 ATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGC 1680

QY 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGAGAGTCTGCACAAAACCCGCTGGACA 1740

Db 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGAGAGTCTGCACAAAACCCGCTGGACA 1740

QY 1741 CCTTCTTCCCTTTTGATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCCTATT 1800

Db 1741 CCTTCTTCCCTTTTGATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCCTATT 1800

QY 1801 ATCAGGTATGGGAAGACATGAGTCTGCTGAAGAGCTACAGGAGTTCAAGAAACCCCATGAAA 1860

Db 1801 ATCAGGTATGGGAAGACATGAGTCTGCTGAAGAGCTACAGGAGTTCAAGAAACCCCATGAAA 1860

QY 1861 AGGACATAGTGGAAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAAATGATA 1920

Db 1861 AGGACATAGTGGAAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAAATGATA 1920

QY 1921 CCGTGATGGGATCACACCAAGCTCCTTTGACACGCTTTCCGAAAGTCTTCAAGTAGTG 1980

Db 1921 CCGTGATGGGATCACACCAAGCTCCTTTGACACGCTTTCCGAAAGTCTTCAAGTAGTG 1980

QY 1981 TGGGCTCCCCACCCGTTGTGTACATGCAACCCAGTCCCTCTGACGGGAGAAATTTGTGA 2040

Db 1981 TGGGCTCCCCACCCGTTGTGTACATGCAACCCAGTCCCTCTGACGGGAGAAATTTGTGA 2040

QY 2041 CTGAGATGTGACATTTGGGATTTCCCAT 2068

Db 2041 CTGAGATGTGACATTTGGGATTTCCCAT 2068

RESULT 2

US-09-952-013A-1  
; Sequence 1, Application US/09952013A  
; Patent No. US20020146801A1  
; GENERAL INFORMATION:  
; APPLICANT: GRUMMT, Ingrid  
; APPLICANT: VINGRON, Martin  
; TITLE OF INVENTION: RNA POLYMERASE I TRANSCRIPTION FACTOR TIF-1A  
; FILE REFERENCE: 38485-0007  
; CURRENT APPLICATION NUMBER: US/09/952,013A  
; CURRENT FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: PCT/DE00/00767  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: DE 199 11 992.9  
; PRIOR FILING DATE: 1999-03-17

```
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (40)..(1992)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (721)..(721)
; OTHER INFORMATION: n is a, c, g or t
; NAME/KEY: misc_feature
; LOCATION: (1435)..(1435)
; OTHER INFORMATION: n is a, c, g or t
US-09-952-013A-1

Query Match      97.6%; Score 2018.4; DB 9; Length 2040;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2019; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 47 GCGGTCGCGTTAGTTCGGCCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 106
Db 18 GCGGTCGCGTTAGTTCGGCCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 77

QY 107 TGGCGGCGCTTCGTCCTCTGCAGTTAAGAAAGCTGGGCGCGTCGAGGACTGGGATTTCAA 166
Db 78 TGGCGGCGCTTCGTCCTCTGCAGTTAAGAAAGCTGGGCGCGTCGAGGACTGGGATTTCAA 137

QY 167 TATGCGTGCAATTAGAGAAATGACTTTTTCAAATCTCCCCCAAGAAAACTGTTGCGTTGG 226
Db 138 TATGCGTGCAATTAGAGAAATGACTTTTTCAAATCTCCCCCAAGAAAACTGTTGCGTTGG 197

QY 227 TGGAACTGTGACAGAAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTT 286
Db 198 TGGAACTGTGACAGAAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTT 257

QY 287 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCCAGATCATCAACTGGTGCT 346
Db 258 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCCAGATCATCAACTGGTGCT 317

QY 347 AGAATTCGGTCTCTCTATCATGTACTTGACAAAGACTTTGAGCAACTTATCAGTATTAT 406
Db 318 AGAATTCGGTCTCTCTATCATGTACTTGACAAAGACTTTGAGCAACTTATCAGTATTAT 377

QY 407 ATTAAGATTGCCTTGGTTGAATAGAAAGTCAACACAGTAGTGGAAAGATTTGGCTTTTCT 466
Db 378 ATTAAGATTGCCTTGGTTGAATAGAAAGTCAACACAGTAGTGGAAAGATTTGGCTTTTCT 437

QY 467 TGGTAATCTTGATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 526
Db 438 TGGTAATCTTGATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 497

QY 527 CCATTTTGTGCCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 586
Db 498 CCATTTTGTGCCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 557

QY 587 TGATGAAGATGATAATCTTCCTGCAAAATTTTGACACATGTACAGAGCCCTTGCAAAATA 646
Db 558 TGATGAAGATGATAATCTTCCTGCAAAATTTTGACACATGTACAGAGCCCTTGCAAAATA 617

QY 647 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAAAATTTCC 706
Db 618 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAAAATTTCC 677

QY 707 ATTTGTTCCGAAATCAGAGAAACACTGGAAATGTTACGTTCAATACTTAAAGGATTAG 766
Db 678 ATTTGTTCCGAAATCAGAGAAACACTGGAAATGTTACGTTCAATACTTAAAGGATTAG 737

QY 767 TGTATATTTTCCAAACCTTGAGGCATGAAATTTCTGGAGCTTATTTATGAAAACTACTCAA 826
Db 738 TGTATATTTTCCAAACCTTGAGGCATGAAATTTCTGGAGCTTATTTATGAAAACTACTCAA 797
```

```
QY 827 GTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 886
Db 798 GTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 857

QY 887 TTTGGTGGGACAGATTCCACGGAAGGATTGTTTAAATATGGATGAAGAAACTGA 946
Db 858 TTTGGTGGGACAGATTCCACGGAAGGATTGTTTAAATATGGATGAAGAAACTGA 917

QY 947 ACATGAAAAAAGGGTCTCTGAAACGGCTCGACAGATGGTGCATCCTGTAGCCGAGCG 1006
Db 918 ACATGAAAAAAGGGTCTCTGAAACGGCTCGACAGATGGTGCATCCTGTAGCCGAGCG 977

QY 1007 CCTGGACATCCTGATGTCCTTTGGTTTGTCTACATGAAGGATGTCTGCTATGTAGTGG 1066
Db 978 CCTGGACATCCTGATGTCCTTTGGTTTGTCTACATGAAGGATGTCTGCTATGTAGTGG 1037

QY 1067 TAAGGTTGATAACGGCAAAAAAAGGATCTATATCGCGACCTGTATAAACATCTTTGACAA 1126
Db 1038 TAAGGTTGATAACGGCAAAAAAAGGATCTATATCGCGACCTGTATAAACATCTTTGACAA 1097

QY 1127 ACTCCTGTTGCCACCCATGCCCTCCTGCCATGTACAGTCTTTTCACTGTTTACCTCTGTAG 1186
Db 1098 ACTCCTGTTGCCACCCATGCCCTCCTGCCATGTACAGTCTTTTCACTGTTTACCTCTGTAG 1157

QY 1187 TTTCAAATTTGGGATTCGCAGAGGCAATTTTGGAAACATCTCTGAAAAAATTCAGGACCC 1246
Db 1158 TTTCAAATTTGGGATTCGCAGAGGCAATTTTGGAAACATCTCTGAAAAAATTCAGGACCC 1217

QY 1247 AAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATATATTGGAAGCTTTTGGCAAG 1306
Db 1218 AAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATATATTGGAAGCTTTTGGCAAG 1277

QY 1307 AGCTAAATTTATTCCTTATTACTGTAAAAATCATGCTAGATCTTTTGGTTAACTGGCT 1366
Db 1278 AGCTAAATTTATTCCTTATTACTGTAAAAATCATGCTAGATCTTTTGGTTAACTGGCT 1337

QY 1367 GCACATATACCTTAATAACAGGATTCGGGAACAAAGSCATTTCTGCGATGTTGCTCTCCA 1426
Db 1338 GCACATATACCTTAATAACAGGATTCGGGAACAAAGSCATTTCTGCGATGTTGCTCTCCA 1397

QY 1427 TGGACCAATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCA 1486
Db 1398 TGGACCAATTTACTCAGCCTGCCAAGCTGTGTTCTACNCCTTTGTTTTAGACACAAGCA 1457

QY 1487 GCTTTTGAGCGGAAACCTGAAAGAGGTTTGCAATCTTTCAGAGTCTGAAATTTTGAGCG 1546
Db 1458 GCTTTTGAGCGGAAACCTGAAAGAGGTTTGCAATCTTTCAGAGTCTGAAATTTTGAGCG 1517

QY 1547 GATAGTGATGAGCCAGCTAAAATCCCCTGAAAGATTTGCGTGCCTCAGTGGTTAACTTTT 1606
Db 1518 GATAGTGATGAGCCAGCTAAAATCCCCTGAAAGATTTGCGTGCCTCAGTGGTTAACTTTT 1577

QY 1607 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA 1666
Db 1578 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA 1637

QY 1667 TCGCCAGATGCTGCCAGTCATTAGGAGTACCCTGCGCTGGAGGAGACTCAGTGCAGATCTGCAC 1726
Db 1638 TCGCCAGATGCTGCCAGTCATTAGGAGTACCCTGCGCTGGAGGAGACTCAGTGCAGATCTGCAC 1697

QY 1727 AAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGCTGAAGAGGTCAAAGAAAT 1786
Db 1698 AAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGCTGAAGAGGTCAAAGAAAT 1757

QY 1787 CATTGATCCTATTTATCAGGTATGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAA 1846
Db 1758 CATTGATCCTATTTATCAGGTATGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAA 1817

QY 1847 GAAACCCATGAAAAAGGACATAGTGGAAAGATGAAGATGATGACTTTTCTGAAAGGCGAAGT 1906
Db 1818 GAAACCCATGAAAAAGGACATAGTGGAAAGATGAAGATGATGACTTTTCTGAAAGGCGAAGT 1877
```

```
QY 1907 GCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAG 1966
|
Db 1878 GCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAG 1937
|
QY 1967 TCCITCAAGTAGTGTGGGCTCCCCACCCGTGTTGTATCATGCAACCCAGTCCCCCTCTGACG 2026
|
Db 1938 TCCITCAAGTAGTGTGGGCTCCCCACCCGTGTTGTATCATGCAACCCAGTCCCCCTCTGACG 1997
|
QY 2027 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 2068
|
Db 1998 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 2039
|
RESULT 3
US-10-094-466-7
; Sequence 7, Application US/10094466
; Publication No. US20030203363A1
; GENERAL INFORMATION:
; APPLICANT: Spytex et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS OF USING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-290D
; CURRENT APPLICATION NUMBER: US/10/094,466
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/288,148
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/338,375
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/275,579
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/335,302
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/275,601
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/276,000
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/277,338
; PRIOR FILING DATE: 2001-03-20
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatIn 2.1
; SEQ ID NO 7
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (40)..(1723)
US-10-094-466-7

Query Match 70.5%; Score 1457.6; DB 16; Length 1770;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 1743; Conservative 0; Mismatches 9; Indels 270; Gaps 1;

QY 47 GGGTCGGTTAGTTCGGCCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCGGGAGA 106
|
Db 18 GGGTCGGTTAGTTCGGCCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCGGGAGA 77
|
QY 107 TGGGCGCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGGCTCGAGGACTGGGATTTCAA 166
|
Db 78 TGGGCGCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGGCTCGAGGACTGGGATTTCAA 137
|
QY 167 TATCGGTGCATTAGAGAATGACTTTTCAATTCTCCCCCAAGAAAAAATGTTCCGTTTGG 226
|
Db 138 TATCGGTGCATTAGAGAATGACTTTTCAATTCTCCCCCAAGAAAAAATGTTCCGTTTGG 197
|
```

```
QY 227 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGCTGAAAAAATGACTTTTGAGTT 286
|
Db 198 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGCTGAAAAAATGACTTTTGAGTT 257
|
QY 287 GTTGAAGAACCCAGCTGTTAGATCCAGACATATAAAGGATGACCAGATCATCAACTGGCTGCT 346
|
Db 258 GTTGAAGAACCCAGCTGTTAGATCCAGACATATAAAGGATGACCAGATCATCAACTGGCTGCT 317
|
QY 347 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTTGAGCAACTTATCAGTATTAT 406
|
Db 318 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTTGAGCAACTTATCAGTATTAT 377
|
QY 407 ATTAAGATTGCCTTGGTTGAATAGAAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCT 466
|
Db 378 ATTAAGATTGCCTTGGTTGAATAGAAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCT 437
|
QY 467 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 526
|
Db 438 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 497
|
QY 527 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTTCAGATTCTGA 586
|
Db 498 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTTCAGATTCTGA 557
|
QY 587 TGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATAAT 646
|
Db 558 TGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATAAT 617
|
QY 647 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 706
|
Db 618 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 677
|
QY 707 ATTTGTTGAAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGATTAG 766
|
Db 678 ATTTGTTGAAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGATTAG 737
|
QY 767 TGTATATTTTCCAACTTTGAGGCATGAAATCTTGAGAGCTTATTATTGAAAAACTACTCAA 826
|
Db 738 TGTATATTTTCCAACTTTGAGGCATGAAATCTTGAGAGCTTATTATTGAAAAACTACTCAA 797
|
QY 827 GTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAAACAGCAACTCAAAC 886
|
Db 798 GCTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAAACAGCAACTCAAAC 857
|
QY 887 TTGTGTTGGGACAGATTCCACGGAAGGATTGTTTAATATGATGAAGATGAAGAAACTGA 946
|
Db 858 TTGTGTTGGGACAGATTCCACGGAAGGATTGTTTAATA----- 895
|
QY 947 ACATGAAAACAAAGGCTGGTCTTGAACGGCTCGACACAGATGGTGCATCCTGTAGCCGACG 1006
|
Db 896 ----- 895
|
QY 1007 CCTGGACATCCTGATGTCCTTTGGTTTGTCTCCTACATGAAGGATGCTGCTATGTAGATGG 1066
|
Db 896 ----- 895
|
QY 1067 TAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGTATAAACATCTTTGACAA 1126
|
Db 896 ----- 895
|
QY 1127 ACTCCTGTTGCCACCCATGCCCTCCTGCCATGTACAGTTTTTTCATGTTTTACCTCTGTAG 1186
|
Db 896 ----- 895
|
QY 1187 TTTCAAATTTGGGATTCGCAGAGGCAATTTTGGAAACATCTCTGGAAAAAATTCAGAGACCC 1246
|
Db 896 -----TGGGATTCGCAGAGGCAATTTTGGAAACATCTTTGGAAAAAATTCAGAGATCC 947
|
QY 1247 AAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATATATTTGGAAGCTTTTTTGGCAAG 1306
|
Db 948 AAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATATATTTGGAAGCTTTTTTGGCAAG 1007
|
QY 1307 AGCTAAATTTAFTCCTCTTATTACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCT 1366
|
```

Db 1008 AGCTAAATTATTCTCTATTACTGTAAACCATGCCCTAGATCTTTGGTTAACTGGCT 1067  
QY 1367 GCACATATACCTTAATAACCAAGGATTCGGGAACAAAGGATTCGCGATGTTGCTCTCA 1426  
Db 1068 GCACATATACCTTAATAACCAAGGATTCGGGAACAAAGGATTCGCGATGTTGCTCTCA 1127  
QY 1427 TGGACCAATTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACAAGCA 1486  
Db 1128 TGGACCAATTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACAAGCA 1187  
QY 1487 GCTTTTGAGCGGAAACCTGAAAGAGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG 1546  
Db 1188 GCTTTTGAGCGGAAACCTGAAAGAGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG 1247  
QY 1547 GATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCCTGCCCCTCAGTGGTTAACTTTT 1606  
Db 1248 GATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCCTGCCCCTCAGTGGTTAACTTTT 1307  
QY 1607 TGCTGCAATCAAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA 1666  
Db 1308 TGCTGCAATCAAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA 1367  
QY 1667 TCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCAC 1726  
Db 1368 TCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCAC 1427  
QY 1727 AAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAAGAAAT 1786  
Db 1428 AAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAAGAAAT 1487  
QY 1787 CATTGATCCTATTATCAGGTATGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAA 1846  
Db 1488 CATTGATCCTATTATCAGGTGTGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAA 1547  
QY 1847 GAAACCCATGAANAAGGACATAGTGAAGATGAAGATGAATGACTTTCTGAAAGGCGAAGT 1906  
Db 1548 GAAACCCATGAANAAGGACATAGTGAAGATGAAGATGAATGACTTTCTGAAAGGCGAAGT 1607  
QY 1907 GCCCCAGATGATACCGTGATGGGATCAACCAAGCTCCTTTGACACGCAATTTCCGAAG 1966  
Db 1608 GCCCCAGATGATACCGTGATGGGATCAACCAAGCTCCTTTGACACGCAATTTCCGAAG 1667  
QY 1967 TCCTTCAAGTAGTGTGGGCTCCCCACCCGTTGTACATGCAACCCAGTCCCCTCTGACG 2026  
Db 1668 TCCTTCAAGTAGTGTGGGCTCCCCACCCGTTGTACATGCAACCCAGTCCCCTCTGACG 1727  
QY 2027 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 2068  
Db 1728 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 1769

RESULT 4  
US-10-094-749-544  
; Sequence 544, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHIKO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOYUKI

; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 544  
; LENGTH: 2493  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-094-749-544  
  
Query Match 27.4%; Score 567; DB 16; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 3.9e-160;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1196 GGGATTCCGAGAGGCGATTTTGGAAACATCTCTGGAACAAATTCAGGACCCCAAGTAATCC 1255  
Db 711 GGGATTCCGAGAGGCGATTTTGGAAACATCTTTGGAAAAAATTCAGGATCCAAGTAATCC 770  
QY 1256 TGCCATCATCAGGCGCTGCTGGAATATATTTGGAAGCTTTTGGCAAGAGCTAAAT 1315  
Db 771 TGCCATCATCAGGCGCTGCTGGAATATATTTGGAAGCTTTTGGCAAGAGCTAAAT 830  
QY 1316 TATTCCTCTTATTACTGTAAATCATGCCTAGATCTTTGGTTAACTGGCTGCACATATA 1375  
Db 831 TATTCCTCTTATTACTGTAAACCATGCCTAGATCTTTGGTTAACTGGCTGCACATATA 890  
QY 1376 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTTCTGCCATGTTGCTCTCCATGGACCATT 1435  
Db 891 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTTCTGCCATGTTGCTCTCCATGGACCATT 950  
QY 1436 TTACTAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAAGCAGCTTTTGAG 1495  
Db 951 TTACTAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAAGCAGCTTTTGAG 1010  
QY 1496 CGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 1555  
Db 1011 CGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 1070  
QY 1556 GAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTGTGCAAT 1615  
Db 1071 GAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTGTGCAAT 1130  
QY 1616 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTAGAGAGGAACAATCGCCAGAT 1675  
Db 1131 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTCGAGAGGAACAATCGCCAGAT 1190  
QY 1676 GCTGCCAGTCATTAGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCT 1735  
Db 1191 GCTGCCAGTCATTAGAGTACCGCTGGAGGAGACTCAGTGCAGACCTGCAACAAACCCACT 1250  
QY 1736 GGACACCTTCTTCCCCTTTGATCCCTGTGCTGAAGAGGTCA 1778  
Db 1251 GGACACCTTCTTCCCCTTTGATCCCTGTGCTGAAGAGGTGA 1293

RESULT 5  
US-10-625-172  
; Sequence 172, Application US/10221625  
; Publication No. US20040033942A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: HILLMAN, Jennifer L.  
; APPLICANT: BAUGHN, Mariah R.  
; APPLICANT: YUE, Henry  
; APPLICANT: LAL, Preeti





```
QY 1343 CCTAGATCTTTGGTTAACTGGCTGCACATAPACCTTAATAACCAAGGATTCGGGAACAAA 1402
Db 207 CCTAGATCTTTGGTTAACTGGCTGCACATAPACCTTAATAACCAAGGATTCGGGAACAAA 148
QY 1403 GGCATTCTGGCATGTTGCTCTCCATGGACCACTTTACTCAGCTGCCAAGCTGTGTTCTA 1462
Db 147 GGCATTCTGGCATGTTGCTCTCCATGGACCACTTTACTCAGCTGCCAAGCTGTGTTCTA 88
QY 1463 CACCTTTGTTTTAGACACAAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGT 1514
Db 87 CACCTTTGTTTTAGACACAAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGT 36

RESULT 8
US-10-029-386-12002
; Sequence 12002, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12002
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; OTHER INFORMATION: NT HIT: gil16164646, EVALUE 1.00e-100
; OTHER INFORMATION: SWISSPROT HIT: P48322, EVALUE 3.00e-02
; OTHER INFORMATION: EST_HUMAN HIT: AW768543.1, EVALUE 1.00e-100
US-10-029-386-12002

Query Match 9.1%; Score 188.8; DB 15; Length 535;
Best Local Similarity 88.4%; Pred. No. 6.5e-46;
Matches 205; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1283 TTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTTATTACTGTAAATCATG 1342
Db 269 TTTTATTGAAGTGTAAGCTTGACTTGAATTTTCTTTCTCTAGTACTGTAAACCATG 328
QY 1343 CCTAGATCTTTGGTTAACTGGCTGCACATATACCTTAATAACCAAGGATTCGGGAACAAA 1402
Db 329 CCTAGATCTTTGGTTAACTGGCTGCACATATACCTTAATAACCAAGGATTCGGGAACAAA 388
QY 1403 GGCATTCTGGCATGTTGCTCTCCATGGACCACTTTACTCAGCCTGCCAAGCTGTGTTCTA 1462
Db 389 GGCATTCTGGCATGTTGCTCTCCATGGACCACTTTACTCAGCCTGCCAAGCTGTGTTCTA 448
QY 1463 CACCTTTGTTTTAGACACAAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGT 1514
Db 449 CACCTTTGTTTTAGACACAAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGT 500

RESULT 9
US-10-029-386-24313/c
; Sequence 24313, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
```

```
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24313
; LENGTH: 184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P48322, EVALUE 1.10e-02
; OTHER INFORMATION: NT HIT: gil4779157, EVALUE 2.00e-99
; OTHER INFORMATION: EST_HUMAN HIT: AW408066.1, EVALUE 3.00e-99
US-10-029-386-24313
```

```
Query Match 8.9%; Score 184; DB 15; Length 184;
Best Local Similarity 100.0%; Pred. No. 8.3e-45;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1328 TACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 1387
Db 184 TACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 125
QY 1388 GGATTCCGGGAACAAAGGCATTCTCGATGTTGCTCTCCATGGACCACTTTACTCAGCCTG 1447
Db 124 GGATTCCGGGAACAAAGGCATTCTCGATGTTGCTCTCCATGGACCACTTTACTCAGCCTG 65
QY 1448 CCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAA 1507
Db 64 CCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAA 5
QY 1508 AGAA 1511
Db 4 AGAA 1
```

```
RESULT 10
US-10-029-386-25702
; Sequence 25702, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25702
; LENGTH: 184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; OTHER INFORMATION: SWISSPROT HIT: P48322, EVALUE 1.50e-02
; OTHER INFORMATION: NT HIT: gil16164646, EVALUE 2.00e-99
; OTHER INFORMATION: EST_HUMAN HIT: AI660308.1, EVALUE 3.00e-99
US-10-029-386-25702
```

```
Query Match 8.8%; Score 182.4; DB 15; Length 184;
Best Local Similarity 99.5%; Pred. No. 2.5e-44;
Matches 183; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1329 ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 1388
Db 1 ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 60
```

```

QY 1389 GATTCGGGAACAAAGGCAATTCGCGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGC 1448
Db 61 GATTCGGGAACAAAGGCAATTCGCGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGC 120
QY 1449 CAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGACGGGAAACCTGAAA 1508
Db 121 CAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGACGGGAAACCTGAAA 180
QY 1509 GAAG 1512
Db 181 GAAG 184

RESULT 11
US-09-998-598-1177/c
; Sequence 1177, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1177
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1177

Query Match 7.5%; Score 156; DB 9; Length 557;
Best Local Similarity 100.0%; Pred. No. 6e-36;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1913 GAATGATACCGTGATGGGATCACAACCAAGCTCCTTTGACACGCAATTTCCGAAAGTCCTTC 1972
Db 557 GAATGATACCGTGATGGGATCACAACCAAGCTCCTTTGACACGCAATTTCCGAAAGTCCTTC 498
QY 1973 AAGTAGTGTGGCTCCCCACCCGCTGTGTACATGCAACCCAGTCCCCTCTGACGGCAGAA 2032
Db 497 AAGTAGTGTGGCTCCCCACCCGCTGTGTACATGCAACCCAGTCCCCTCTGACGGCAGAA 438
QY 2033 ATTTGTGACTGAGATGTGACATTTGGGATTCCCCAT 2068
Db 437 ATTTGTGACTGAGATGTGACATTTGGGATTCCCCAT 402

RESULT 12
US-10-029-386-7910/c
; Sequence 7910, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7910
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (153)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (154)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (161)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (178)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (187)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (203)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (209)
; OTHER INFORMATION: n equals a,t,g, or c

```

	Query Match	7.1%;	Score 146.6;	DB 9;	Length 225;
	Best Local Similarity	91.9%;	Pred. No. 2.1e-33;		
	Matches 171; Conservative	0;	Mismatches 13;	Indels 2;	Gaps 2;
Qy	851	GGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACCTGTGCTGGGACAGATTCCACCGGA	910		
Dd	8	GTGTATTTAAGATGCTGAAGAAACAGCAACTCAAACCTNGTNGTGGACAGATTCCACCGGA	67		
Qy	911	AGGATTGTTTTAATATGGATGAAGATG-AAGRAACTGAACATGAACAACAAAGCCTGGTCCTG	969		
Dd	68	AGGANTGNTTAATATGGATGAAGATGNAAGRAACTGAACATGAACAACAAAGCCTGGTCCTG	127		
Qy	970	AACGGCTCGACCAGATGGTGATCCTGTAGCGGAGCGCCT-GGACATCCTGATGTCTTTG	1028		
Dd	128	AACGGCTTACCAGATGGTGCATCCNNTAGCCGCGCCCTGGGACATCCGNATGTCTTTN	187		
Qy	1029	GTTTTG 1034			
Dd	188	GTTTTG 193			

```

RESULT 14
US-10-029-386-8224
; Sequence 8224, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8224
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q9NYV4, EVALUE 7.10e-01
; OTHER INFORMATION: EST HUMAN HIT: BG485267.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: gil16164646, EVALUE 5.00e-84
US-10-029-386-8224

```

	Query Match	6.8%;	Score 140.6;	DB 15;	Length 506;
	Best Local Similarity	96.9%;	Pred. No. 2.6e-31;		
	Matches 154;	Conservative 0;	Mismatches 4;	Indels 1;	Gaps 1;
QY	1	ACAGAGGCTGTGGCTGGAAGAGACTGGGCATCCGGCCTCAGCGGCAGCGGTTCGCGTTAGT	60		
Db	214	ACAGAGGCTGTGGCTGGAAGAGACTGGGCATCCGGCCTCAGCGGCAGCGGTTCGCGTTAGT	273		
QY	61	TGGGCCCAATGGCGGCACCGCTGTCTTCACACG - CGTTTGCGGGAGATGCGGCGCGCTTCG	119		
Db	274	TGGGCCCAATGGCGGCACCGCTGTCTTCACAGTTGTTTGTCGGGAGATGCGGCGCGCTTCG	333		
QY	120	TCCTCTGCAGTTAAGAAGCTGGGCGCGTCTGAGGACTGGG	158		
Db	334	TCCTCTGCAGTCAAGACGTGGGCGCGTCTGAGGACTGGG	372		

```

RESULT 15
US-10-029-386-21924
; Sequence 21924, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21924
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P09133, EVALUAE 5.10e-01
; OTHER INFORMATION: NT HIT: g16164646, EVALUAE 1.00e-83
; OTHER INFORMATION: EST_HUMAN HIT: BG485267.1, EVALUAE 3.00e-76
US-10-029-386-21924

```

	Query Match	6.8%;	Score 139.6;	DB 15;	Length 342;
	Best Local Similarity	96.8%;	Pred. No. 3.9e-31;		
	Matches 153;	Conservative	0;	Mismatches 4;	Indels 1; Gaps 1
Qy	1	ACAGAGGCTGTGGCTGGAAAGGAGCTGGGCATCCGGCCTAGGCGCAGCGTCCGCTTAGT	60		
Db	185	ACAGAGGCTGTGGCTGGAAAGGAGCTGGGCATCCGGCCTAGGCGCAGCGTCCGCTTAGT	244		
Qy	61	TCGGCCCCAATGGCGGCACCGCTGCTTCACACG-CGTTTCCGGGAGATGCGGCCGCTTCG	119		
Db	245	TCGGCCCCAATGGCGGCACCGCTGCTTCACACGTTGTTGTGGGAGATGCGGCCGCTTCG	304		
Qy	120	TCCTCTGCAGTTAAGAAAGCTGGGCGCGTTCGAGGACTGG	157		
Db	305	TCCTCTGCAGTCAAGACGCTGGGCGCGTCGAGGACTGG	342		

Search completed: July 18, 2004, 08:45:45  
Job time : 937 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:56:33 ; Search time 772 Seconds  
(without alignments)  
4113.969 Million cell updates/sec

Title: US-09-932-678-2  
Perfect score: 3403  
Sequence: 1 MAAPLLHTRLPGDAAASSA.....RSPSSSVGSPPLVLMQPSPL 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09932678/runat\_15072004\_103606\_5312/app\_query.fasta\_1.839  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosu62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09932678 @CGN 1 1 480 @runat 15072004 103606\_5312  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	3403	100.0	2068	9	US-09-932-678-1	Sequence 1, Appli
2	3391	99.6	2040	9	US-09-952-013A-1	Sequence 1, Appli
3	2839	83.4	1770	16	US-10-094-466-7	Sequence 7, Appli
4	1124	33.0	1461	13	US-10-221-625-172	Sequence 172, App
5	1000	29.4	2493	16	US-10-094-749-544	Sequence 544, App
6	565.5	16.6	3878	15	US-10-128-714-207	Sequence 207, App
7	565.5	16.6	4139	15	US-10-128-714-5207	Sequence 5207, Ap
8	563	16.5	2139	15	US-10-128-714-6207	Sequence 6207, Ap
9	563	16.5	2139	15	US-10-128-714-7207	Sequence 7207, Ap
10	561.5	16.5	1878	15	US-10-128-714-1207	Sequence 1207, Ap
11	542.5	15.9	1830	15	US-10-128-714-2207	Sequence 2207, Ap
12	506.5	14.9	2702	13	US-10-424-599-77428	Sequence 77428, A
13	499.5	14.7	1671	15	US-10-032-585-6770	Sequence 6770, Ap
14	456.5	13.4	1815	9	US-09-938-842A-2204	Sequence 2204, Ap
15	456.5	13.4	1815	11	US-09-938-842A-2204	Sequence 2204, Ap
16	391.5	11.5	1293	16	US-10-320-797-2207	Sequence 2207, Ap
17	375	11.0	3902	16	US-10-320-797-207	Sequence 207, App
c 18	341	10.0	535	15	US-10-029-386-10613	Sequence 10613, A
19	337.5	9.9	1902	16	US-10-320-797-1207	Sequence 1207, Ap
20	335	9.8	1284	13	US-10-425-114-36547	Sequence 36547, A
21	333	9.8	535	15	US-10-029-386-12002	Sequence 12002, A
22	332.5	9.8	1103	17	US-10-437-963-99865	Sequence 99865, A
c 23	332	9.8	184	15	US-10-029-386-24313	Sequence 24313, A
24	327	9.6	184	15	US-10-029-386-25702	Sequence 25702, A
25	322	9.5	2103	9	US-09-925-301-192	Sequence 192, App
26	265.5	7.8	2591	17	US-10-437-963-99864	Sequence 99864, A
27	207	6.1	225	9	US-09-783-590-11086	Sequence 11086, A
28	202	5.9	140	9	US-09-728-445-420	Sequence 420, App
c 29	189	5.6	557	9	US-09-998-598-1177	Sequence 1177, Ap
30	180	5.3	414	9	US-09-864-761-4848	Sequence 4848, Ap
31	161	4.7	95	9	US-09-864-761-21583	Sequence 21583, A
c 32	135	4.0	598	15	US-10-029-386-7910	Sequence 7910, Ap
33	132.5	3.9	7884	13	US-10-424-599-113868	Sequence 113868, A
34	130.5	3.8	3255	13	US-10-424-599-754	Sequence 754, App
c 35	128	3.8	89	15	US-10-029-386-21610	Sequence 21610, A
36	127	3.7	1908	15	US-10-032-585-6636	Sequence 6636, Ap
37	120.5	3.5	2841	13	US-10-282-122A-18265	Sequence 18265, A
38	118	3.5	10116	17	US-10-437-963-98312	Sequence 98312, A
39	117.5	3.5	3018	13	US-10-425-114-12808	Sequence 12808, A
40	117	3.4	2628	15	US-10-032-585-6415	Sequence 6415, Ap
41	116.5	3.4	4371	17	US-10-437-963-92083	Sequence 92083, A
42	116	3.4	2483	13	US-10-424-599-12638	Sequence 12638, A
43	116	3.4	5164	17	US-10-437-963-14824	Sequence 14824, A
44	115	3.4	4398	13	US-10-282-122A-11087	Sequence 11087, A
45	112	3.3	506	15	US-10-029-386-8224	Sequence 8224, Ap

ALIGNMENTS

RESULT 1  
US-09-932-678-1  
; Sequence 1, Application US/09932678  
; Patent No. US20020090706A1  
; GENERAL INFORMATION:  
; APPLICANT: Reeder, Ronald H.  
; APPLICANT: Moorefield, Beth  
; APPLICANT: Greene, Elizabeth A.  
; TITLE OF INVENTION: HUMAN RRM3 AND COMPOSITIONS AND METHODS RELATING  
; TITLE OF INVENTION: THERETO  
; FILE REFERENCE: 14538A-005810US  
; CURRENT APPLICATION NUMBER: US/09/932,678  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: 60/225,893  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2068  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-932-678-1

Alignment Scores:		0	Length:	2068
Pred. No.:	Score:	3403.00	Matches:	651
Percent Similarity:		100.00%	Conservative:	0
Best Local Similarity:		100.00%	Mismatches:	0
Query Match:		100.00%	Indels:	0
DB:		9	Gaps:	0
US-09-932-678-2 (1-651) x US-09-932-678-1 (1-2068)				
QY	1	MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaLaSerSerSerAla	20	
Db	69	ATGGCGCACCGCTGCTTCACACGCGTTTGCCGGGAGATCGGCCGCTTCGTCTCTGCA	128	
QY	21	VallLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp	40	
Db	129	GTTAAGAAGCTGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATAGAGAAATGAC	188	
QY	41	PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu	60	
Db	189	TTTTTCAATTCTCCCCAAGAAAACGTGTCGGTTTGGTGGAACTGTGACAGAAAGTCTTG	248	
QY	61	LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp	80	
Db	249	CTGAAGTACAAAAGGTTGAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT	308	
QY	81	ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet	100	
Db	309	CCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAAATTCGGTCTTCTATCATG	368	
QY	101	TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn	120	
Db	369	TACTTGACAAAAGACTTTGAGCAACTATCAGTATATATTAAGATTGCCCTTGGTTGAAT	428	
QY	121	ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln	140	
Db	429	AGAACTCAAAACAGTAGTGAAGAGATATTGGCTTTTCTTGGTAATCTTGATCAGCACAG	488	
QY	141	ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal	160	
Db	489	ACTGTTTTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCGAGTG	548	
QY	161	IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro	180	
Db	549	ATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATACCATCGACA	608	
QY	181	AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr	200	
Db	609	GCAAAATTTTGACACATGTACAGAGCCCTTGCAAAATAATAGCAAGATATGTACCATCGACA	668	
QY	201	ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg	220	
Db	669	CCGTGGTTTCTCATGCGCAATACTGGTGGAAAAATTTCCATTGTTCCGAAAAATCAGAGAGA	728	
QY	221	ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg	240	
Db	729	ACACTGGAAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTTCCAAACCTTGAGG	788	
QY	241	HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg	260	
Db	789	CATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTGAATGCATCCCGG	848	
QY	261	GlnGlyIleLeuGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr	280	
Db	849	CAGGTTATTGAAGATGCTGAAGAAAACAGCAACTCAAACTTGTGTTGGGACAGATTCACAG	908	
QY	281	GluGlyLeuPheAsnMetAspGluGluGluThrGluHisGluThrLysAlaGlyPro	300	
Db	909	GAAGGATTGTTTAAATATGGATGAAGATGAAGAAACTGAACATGAACAAAGGCTGTCTCT	968	
QY	301	GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu	320	
Db	969	GAACGGCTCGACCAAGATGGTGCATCCTGTAGCCGAGCGCTTGACATCCTGTATGTCCTTG	1028	

QY	321	ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr	340	
Db	1029	GTTTTGTCTACATGAAGGATGTCTGCTATGTAGATGGTAAGTTGATAACGGCAAAACA	1088	
QY	341	LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla	360	
Db	1089	AAGGATCTATATCGGACCTGATAAACATCTTTGACAAACTCCTGTGTGCCACCCATGCC	1148	
QY	361	SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu	380	
Db	1149	TCCTGCCATGTACAGTTTTCATGTTTACCTCTCTAGTTTCAAATTTGGGATTTCGCAGAG	1208	
QY	381	AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg	400	
Db	1209	GCATTTTTGGAAACATCTCTGAAAAAATTCAGGACCCCAAGTAATCCTGCCATCATCAGG	1268	
QY	401	GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle	420	
Db	1269	CAGGCTGCTGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTTATCTCTTATT	1328	
QY	421	ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln	440	
Db	1329	ACTGTAAAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAG	1388	
QY	441	AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys	460	
Db	1389	GATTCGGGAACAAAGGCATTTCTGCGATGTGCTCTCCATGGACCATTTTACTCAGCCTGC	1448	
QY	461	GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys	480	
Db	1449	CAAGCTGTGTTCTACACCTTTGTTTTTAGACACACAGCAGCTTTTGGCGGAAACCTGAAA	1508	
QY	481	GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn	500	
Db	1509	GAAGTTTGCAGTATCTTCAGAGTCTGAATTTTGGCGGATAGTGATGAGCCAGCTAAAT	1568	
QY	501	ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr	520	
Db	1569	CCCCTGAAGATTTGCTGCCCTCAGTGGTTAACTTTTGTGCTGCAATCACAAATAAGTAC	1628	
QY	521	GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle	540	
Db	1629	CAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCATT	1688	
QY	541	ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe	560	
Db	1689	AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCCGCTGGACACCTTCTTC	1748	
QY	561	ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal	580	
Db	1749	CCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTA	1808	
QY	581	TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle	600	
Db	1809	TGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAAGAAACCCCATGAAAAAGGACATA	1868	
QY	601	ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle	620	
Db	1869	GTGGAAGATGAAGATGATGACTTTCTGAAAGCGGAGTGCCCCAGAGATGATACCCGTGATT	1928	
QY	621	GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer	640	
Db	1929	GGGATCACACCAAGTCTTTTGACACGCAATTTCCGAAGTCCITTCAGTAGTGTGTGGGTCC	1988	
QY	641	ProProValLeuTyrMetGlnProSerProLeu	651	
Db	1989	CCACCCGTGTTGTACATGCAACCCAGTCCCCCTC	2021	

RESULT 2

US-09-952-013A-1

; Sequence 1, Application US/09952013A

; Patent No. US20020146801A1

; GENERAL INFORMATION:
; APPLICANT: GRUMMT, Ingrid
; APPLICANT: VINGRON, Martin
; TITLE OF INVENTION: RNA POLYMERASE I TRANSCRIPTION FACTOR TIF-1A
; FILE REFERENCE: 38485-0007
; CURRENT APPLICATION NUMBER: US/09/952,013A
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/DE00/00767
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: DE 199 11 992.9
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (40)..(1992)
; OTHER INFORMATION:
; NAME/KEY: misc\_feature
; LOCATION: (721)..(721)
; OTHER INFORMATION: n is a, c, g or t
; NAME/KEY: misc\_feature
; LOCATION: (1435)..(1435)
; OTHER INFORMATION: n is a, c, g or t

US-09-952-013A-1
Alignment Scores:
Pred. No.: 0 Length: 2040
Score: 3391.00 Matches: 649
Percent Similarity: 99.69% Conservative: 0
Best Local Similarity: 99.69% Mismatches: 2
Query Match: 99.65% Indels: 0
DB: 9 Gaps: 0

US-09-932-678-2 (1-651) x US-09-952-013A-1 (1-2040)
1 MetAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20
40 ATGGCGGACCGCTGCTTCACACGCGTTTGC CGGAGATGCGGCCGCTTCGTCCTGCA 99
21 ValLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40
100 GTTAAGAGCTGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCTAGAGAAATGAC 159
41 PhePheAsnSerProProArgLysThrValArgPheGlyThrValThrGluValLeu 60
160 TTTTTCGAATCTCCCCAAGAAAAAACTGTTGCGTTTGTGTGAACTGTGACAGAGTCTTG 219
61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80
220 CTGAAGTACAAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT 279
81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
280 CCAGACATAAAGGATGACCATCATCAACTGGCTGCTAGAAATTCGGTCTTCTATCATG 339
101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn 120
340 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATATATTAAGATTGCTTGGTTGGAAT 399
121 ArgSerGlnThrValValGluGluThrAlaPheLeuGlyAsnLeuValSerAlaGln 140
400 AGAAGTCAAAACAGTAGTGAAGAGTATTGGCTTTTCTTGGTAATCTTTGATATCAGCACAG 459
141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160
460 ACTGTTTTCCTCAGACCGGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCCGAGTG 519
161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180

Db 520 ATCATTAGGAAGGGGATGTAGATGTTTCAGATTCTGTATGATGAAGATGATAATCTTCCT 579
Qy 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
Db 580 GCAAATTTTGACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 639
Qy 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220
Db 640 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAAATTTCCATTGTTTCGAAAAATCAGAGAGA 699
Qy 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240
Db 700 AACTGGAAATGTTACGTTTCATNACTTACTAAGGATTAGTGTATATTTTCCAACTTGAGG 759
Qy 241 HisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260
Db 760 CATGAAATCTGGAGCTTATATATTGAAAAAACTACTCAAGTTGGATGTGAATGCATCCCGG 819
Qy 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280
Db 820 CAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACCTGTGTGGGACAGATTCCACG 879
Qy 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300
Db 880 GAAGGATTGTTTAATATGGATGAAGATGAAGAAACCTGAACATGAAACAAAGGCTGGTCCT 939
Qy 301 GluArgLeuAspGlnMetValHisProValAlaGluAlaGluArgLeuAspIleLeuMetSerLeu 320
Db 940 GAACGGCTCGACCCAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCCTGATGTCTTTG 999
Qy 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340
Db 1000 GTTTTGTCTTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACA 1059
Qy 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360
Db 1060 AAGGATCTATATCGCACCTGATAAACATCTTTGACAAACTCCTGTTGCCACCCCATGCC 1119
Qy 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380
Db 1120 TCCTGCCATGTACAGTTTTCATGTTTACCTCTGTAGTTCAAATTTGGGATTTCGCAGAG 1179
Qy 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400
Db 1180 GCATTTTGGAAACATCTCTGAAAAAATTCAGGACCCAAAGTAATCCTGTCATCATCAGG 1239
Qy 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420
Db 1240 CAGGCTGCTGGAAATTAATATGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATT 1299
Qy 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440
Db 1300 ACTGTAAAATCATGCTTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAG 1359
Qy 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460
Db 1360 GATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC 1419
Qy 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480
Db 1420 CAAGCTGTGTTCTACNCCTTTGTTTGTAGACACAAAGCAGCTTTTGAGCGGAAACCTGAAA 1479
Qy 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500
Db 1480 GAAGGTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGTATGAGCCAGCTAAAT 1539
Qy 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520
Db 1540 CCCCTGAAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTGTGCAATCACAAATAAGTAC 1599
Qy 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540
Db 1600 CAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAAACAATCGCCAGATGCTGCCAGTCATT 1659



```
QY 341 LysAspLeuTyrArgAspLeuLeuAsnIlePheAspLysLeuLeuLeuProThrHisAla 360
Db 897 -----
QY 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380
Db 898 -----GGATTGCGAG 909
QY 381 AlapheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleArg 400
Db 910 GCATTTTGGACATCTTTGGAAAACTTGACGATCCAAGTAATCCTGCCATCATCAGG 969
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420
Db 970 CAGGCTGCTGGAAATTATTTGGAAAGCTTTTGGCAAGAGCTAAATTTATTCTCTTATT 1029
QY 421 ThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440
Db 1030 ACTGTAAACCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAG 1089
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460
Db 1090 GATTCGGGAACAAAGGCATCTCGCATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC 1149
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480
Db 1150 CAAGCTGTGTTCTACACCTTTGTTTTTAGACAACAAGCAGCTTTTGAGCGGAAACCTGAAA 1209
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500
Db 1210 GAAGGTTTGAGTATCCTCAGATCTGAATTTGAGCGGATAGTGATGAGCCAGCTAAAT 1269
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520
Db 1270 CCCCTGAAGATTGCTGCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATAAGTAC 1329
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540
Db 1330 CAGCTCGTCTCTGTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCAIT 1389
QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560
Db 1390 AGGAGTACCGTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTC 1449
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580
Db 1450 CCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTG 1509
QY 581 TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
Db 1510 TGGGAAGACATGAGTGCTGAAGAGCTACAGAGTTCAAGAAACCCATGAAAAAGGACATA 1569
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620
Db 1570 GTGGAAGATGAAGATGATGACTTTCTGAAAGCGGAAGTGCCCGCAGATGATACCGTGATT 1629
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer 640
Db 1630 GGGATCACACCAAGCTCCTTTTGACACGCAATTTCCGAAGTCTTCAAGTAGTGTTGGCTCC 1689
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651
Db 1690 CCACCCGTGTGTACATGCAACCCAGTCCCTC 1722
```

RESULT 4

```
US-10-221-625-172
; Sequence 172, Application US/10221625
; Publication No. US20040033942A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
```

```
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: MATHUR, Preete
; APPLICANT: SHAH, Purvi
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: REDDY, Roopa
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; FILE REFERENCE: PF-0761 PCT
; CURRENT APPLICATION NUMBER: US/10/221,625
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PERL Program
; SEQ ID NO 172
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040033942A1 2646274CB1
US-10-221-625-172
```

```
Alignment Scores:
Pred. No.: 2.74e-129 Length: 1461
Score: 1124.00 Matches: 256
Percent Similarity: 55.44% Conservative: 14
Best Local Similarity: 52.57% Mismatches: 52
Query Match: 33.03% Indels: 166
DB: 13 Gaps: 10
```

US-09-932-678-2 (1-651) x US-10-221-625-172 (1-1461)

```
QY 195 ArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPhe 214
Db 263 AAATACAAAAATTAGGCAACCTGGTGGTTCTCATGCCAATACTGGTGGAATAATTTCCATTT 322
QY 215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerVal 234
Db 323 GTTCGAAATCAGAGAGACACTGGAATGTACGTTTCATAACTTACTAAGATTAGTGTA 382
QY 235 TyrPheProThrLeuArgHisGluIleLeuGluLeuIleGluLysLeuLeuLysLeu 254
Db 383 TATTTTCCAAACCTTGAGGCATGAAATCTGGAGCTTATTATTGAAAAAACTACTCAAGCTG 442
QY 255 AspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCys 274
Db 443 GATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAAAATCAAACCTTGT 502
QY 275 GlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHis 294
Db 503 GGTGGGACAGATTCCACGGAAGGATTCTTTAATATG----- 538
QY 295 GluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeu 314
Db 538 ----- 538
QY 315 AspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCystyrValAspGlyLys 334
Db 538 ----- 538
QY 335 ValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeu 354
Db 538 ----- 538
QY 355 LeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPhe 374
Db 538 ----- 538
QY 375 LysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSer 394
```



Db 539 -----GGATTGCGAGGCAATTTTGGAAACATCTTTGGAAAACTTGCAGGATCCAAGT 592

Qy 395 AsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAla 414

Db 593 AATCCTGCCATCATCAGCGAGGCTGCTGGAAATATATATTGGAAGCTTTTGGCAAGAGCT 652

Qy 415 LysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHis 434

Db 653 AAATTATTCTCTATTACTGTAAACCATGCCCTAGATCTTTTGGTTAACTGGCTGCAC 712

Qy 435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454

Db 713 ATATACCTTAATAACGAGGATTCGGGAACAAGGCATCTGCGATGTTGCTCTCCATGGA 772

Qy 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474

Db 773 CCATTTTACTCAGCCTGCCAAGCTGTGTCTACACCTTGTGTTTGTAGACACAAGCAGCT 832

Qy 475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494

Db 833 TTGAGCGGAACCTGAAAGAAGGTTTGCAGTATCCTCAGAGCTCTGAATTTGAGCGGATA 892

Qy 495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla 514

Db 893 GTGATGAGCCAGCTAAATCCCCTGAAGATTGCTGCCCTCAGTGGTTAACTTTTGTCT 952

Qy 515 AlaIleThrAsnLysTyr-GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnAr 534

Db 953 GCAATCACA---AAGATGAAGACTTGTG----- 977

Qy 534 gGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAs 554

Db 978 -----GATATGGATGGTGGTGATGGT-----TGCACAC 1006

Qy 554 nProLeuAspThrPhePheProPheAspProCysValleuLys----- 568

Db 1007 AAT-ATCAATTTATTTTATACCACCTGAACCGTGCATCTCAAAATGGTTAAGATGGTAAGT 1065

Qy 569 ----- 576

Db 1066 TTTATGTTGTGATTTTACCATAATAAAAAAATGTAGAGGGAAAAACAGTCTGCCCTCC 1125

Qy 576 oIleTyrGlnValTrpGlu-----AspMetSerAlaGluG1 588

Db 1126 ACTTTTG---ATATGGGACTGCTAACATCTCCACCTTCGGTCTCCCTTCTGCCCCACA 1182

Qy 588 uLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAspAspPh 608

Db 1183 TCTGGGCAAGCTAAGAAAGCCT----- 1204

Qy 608 eLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer----- 625

Db 1205 -----GCTGCTCTCTCTCTGGCACCCAGCTGGAATTCATACCCCAACAGCCCTA 1254

Qy 626 ----SerPheAspThrHisPheArgSerProSerSerSerValGlySerProProValle 644

Db 1255 GCCTTCCCACCAGACCCACATTTTCATCCCATCCCATCCCATCCCATCC----- 1306

Qy 644 uTyrMetGlnProSerPro 650

Db 1307 -----CATCCCCCATCCC 1318

RESULT 5

US-10-094-749-544  
; Sequence 544, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHIKO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 544  
; LENGTH: 2493  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-094-749-544

Alignment Scores:

Pred. No.: 2.27e-113 Length: 2493  
Score: 1000.00 Matches: 188  
Percent Similarity: 97.45% Conservative: 3  
Best Local Similarity: 95.92% Mismatches: 5  
Query Match: 29.39% Indels: 0  
DB: 16 Gaps: 0

US-09-932-678-2 (1-651) x US-10-094-749-544 (1-2493)

Qy 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393

Db 703 TATCGCAGGGGATTTCGAGAGGCAATTTTGGAAACATCTTTGGAAAAAATTGCAGGATCCA 762

Qy 394 SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413

Db 763 AGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGA 822

Qy 414 AlaLysPheIleProIleuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu 433

Db 823 GCTAAATTTATTCTCTTATTACTGTAAACCATGCCTAGATCTTTTGGTTAACTGGCTG 882

Qy 434 HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453

Db 883 CACATATACCTTAATAACGAGGATTCGGGAACAAGGCATTCGCAATGTTGCTCTCCAT 942

Qy 454 GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473

Db 943 GGACCAATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTGTGTTTGTAGACACAAGCAG 1002

Qy 474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg 493

Db 1003 CTTTTGAGCGGAACCTTGAAAGAAGTTTGCAGTATCTTCAGAGTCTGAAATTTTGGCGG 1062

Qy 494 IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe 513

Db 1063 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTT 1122

Qy 514 AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533

Db 1123 GCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGTGTACACCATCATCGAGAGGAACAAT 1182

Qy 534 ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr 553



```
QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445
Db 2222 TTGACCTC-----TTAGGCACCATCTGAACAACCTCCGTTCTTGACTATGAA 2269
QY 446 AlaPheCys-----AspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462
Db 2270 GCCACCTGTGAGGCGCTGATCTTCGCCGCTACGGTCCTTTCTACTCCACGGCACAGGCA 2329
QY 463 ValPheTyrThrPheValPheArgHisLysGlnLeu-----LeuSerGlyAsn 478
Db 2330 TTGTGTATATCTTCTGCTTCGCGTGCGAGACCTGACGACCGCAGCGATGGAGGCGGAT 2389
QY 479 LeuLysGluGlyLeuGlnTyrLeu-----GlnSerLeuAsnPhe----- 491
Db 2390 ACCCCTGAACAAATTTGATGAACCTGGAGCCAGAGACATCACGTTCCCTCCCTCGGTCAAG 2449
QY 492 -----GluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508
Db 2450 GAGGTCTTCACCAAGCAATCCATTCCAAACTGAACCCCTCAAAGTATGCTCGCCGCG 2509
QY 509 ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528
Db 2510 ATTGTGTCTGAATTCGCCGCTATGGCTCAGCACCTCAACTTCATGTCGTGTTTCAGCAT 2569
QY 529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSer----- 542
Db 2570 CTCGAGACGAACAAGACTGCGCGTCTCCTCGTCCGAGCATCTCAGCCATGGCAGAC 2629
QY 543 -----ThrAlaGlyGlyAspSerValGlnIleCysThr 553
Db 2630 CCGCGCTTCAGCCATGTCGAGCGAGAAACTCGTGCCGCGGATGACCTGGGTTAT----- 2683
QY 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPhe 573
Db 2684 ---CAGCTTGATGCCTATTTCCTTTGACCCATACCACTACCTCGCAGCCCGCTTGG 2740
QY 574 IleAspProIleTyrGlnValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLys 593
Db 2741 CTGGAGGATGATTACGTCACCTGGCGTGGTATCCCTGGCGTGGACGACGAGGAC----- 2794
QY 594 LysProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluVal 613
Db 2795 -----GAGGCGGATTGGACAGAGTACCTCGACGGGAGCGAA 2830
QY 614 ProGlnAsnAspThrValIleGlyIleThrProSerSerPheAsp-ThrHisPheArgSe 633
Db 2831 TCCGAAGGAGAC--CTGAGCGAGCTCACTGAGACCGACGAAAGACGACTGAATACTCTGA 2887
QY 633 rProSerSerSerVal 638
Db 2888 ACCAAAAACCCCTGTT 2903
RESULT 7
US-10-128-714-5207
; Sequence 5207, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
```

```
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5207
; LENGTH: 4139
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5207
Alignment Scores:
Pred. No.: 1.18e-58 Length: 4139
Score: 565.50 Matches: 182
Percent Similarity: 44.08% Conservative: 138
Best Local Similarity: 25.07% Mismatches: 264
Query Match: 16.62% Indels: 142
DB: 15 Gaps: 24
US-09-932-678-2 (1-651) x US-10-128-714-5207 (1-4139)
QY 13 AspAlaAlaAlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSer 32
Db 1115 AACAGCGCGCAGCTCCCGGGCTAAACACACAAGTTGGCCGACATGGATCTTTCGCTGTCA 1174
QY 33 AsnMetArgAlaLeuGluAsnAspPheAsnSer-----ProProArgLysThrVal 50
Db 1175 CCCTCGCTCGGTCCGTTCCGATGGCAGCGATGGCAGCCCGCAGTCCCAAGGAAGCGGGTT 1234
QY 51 Arg-----PheGlyGlyThrValThr 57
Db 1235 CGGGTTCAATTTGATAAGGATGTGGAATGAGAGAAGCGTCTTGGAGTGGTCAAGAGAA 1294
QY 58 GluValLeuLeuLysTyrLysLysGly----- 66
Db 1295 GAAATGGCTGTGAGTACGGAAGAGAGCGCGCGTGGTTCGTGAAGAACTGCGCAGAGCT 1354
QY 67 -----GluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 78
Db 1355 ATCCAGCGGCATGTCTCGGAACGGATAGTAGAGCCTAGCATCGTATCAAAAGAGATTTTC 1414
QY 79 ---LeuAspProAspIleLysAspAspGln----- 87
Db 1415 TCGGCAGATCCTCGCGCTCTGGACGAAGATGGCTTGCCTCCCGAAGATTTACCAACGCAC 1474
QY 88 -----IleIleAsnTrpLeuGluPheArgSerSerIleMetTyrLeuThrLysAsp 105
Db 1475 ACTACATTGAAGCATCATCTCATGGGTCTTCTGTCCAATGTTGCGTTCGTGGATCGCAAT 1534
QY 106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrVal 125
Db 1535 TGCAACGGCTTGGTCAATGCAGTCCCTGCACAGTGAATGGCTCGGCGGGATGAGTCGTAT 1594
QY 126 ValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145
Db 1595 ATCAAGTTGTACATTGATTTCTAGGTAACCTTGGCAGCCAGGAGGAAAGCTATCTCGGG 1654
QY 146 ProCysLeuSerMetIleAlaSerHisPhe-----Val 156
Db 1655 GCGGTCTTGAAGATTTGGCCAACAATTTCCGGGAGATCCCAAAGGGACGGGCAAGCTA 1714
QY 157 ProProArgValIleIleLysGluGlyAspValAspValSerAspSerAspGluAsp 176
Db 1715 CTGGCTACGCCCCAGTCCCAAGCTGCAGAGATC----- 1747
QY 177 AspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyr 196
Db 1748 -----TATAGCAGAACTCATATGGCTCTACGTATGATGATGATG 1789
QY 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPhePropheValArg 216
Db 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPhePropheValArg 216
```

DB	1790	ATTCTTCTGGTAGTGGGACCCCTATATCCCCCATCTTGTCTATGCAGTTTCCCTTCGATACG	1841
QY	217	LysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe	236
DB	1850	GATTCCGCAAGGCCAACATTGCGGTACACGGCGAATCTTCTAAAGGTCATTAGTTACGCC	1909
QY	237	ProThrLeuArgHisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspVal	256
DB	1910	CCTGAGCTCCAAGCCGACATTCTAGCTCTCATCACGGAGAAAGTCGTGAAGATCGATGTC	1969
QY	257	AsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGly	276
DB	1970	CAGATT--CAGGTAGATATGGAAGACATCAGGAGCAGAGTGGCGAGGATGTTCTCCAC	2026
QY	277	ThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThr	296
DB	2027	GCAGTCTCGCCTGAAGCCACGATGTTGGAGGATGACGAGAATGAAGTTGACAATGCATCC	2086
QY	297	LysAlaGlyProGluArgLeuAsp-----GlnMetValHisProValAlaGlu--	312
DB	2087	GTCGCTAGCGATGATTCCTTAGACGTGGAGTCCCGCGGAGTCAAGGCCATCAAAGACAAC	2146
QY	313	-----ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCystyr	330
DB	2147	ATACTCAATGGATGGCATGATCGATTTCGTGTTTGAATACTACGCCCGCGCT---TTC	2203
QY	331	ValAspGlyLysValAspAsnGlyLysThrLys--AspLeuTyrArgAspLeuIleAsn	349
DB	2204	ACCACGGGCATTTGGATGACAAGAGAGATGCACTGGATTTA-----CTTCTGAGC	2254
QY	350	IlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPhe	369
DB	2255	CATTTCCAAAGTATTATCCTTCCGACCTATCGCTCCCGCACCTCTCAATTCTCTCTTC	2314
QY	370	TyrLeuCysSerPheLysLeu-----GlyPheAlaGluAlaPheLeuGluHis	385
DB	2315	CACTTCTCCAGTCGTGGCTATTCTGTGGATCGATTGCGCGCAACCTGTGTTTCAGCTC	2374
QY	386	LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn	405
DB	2375	ATCTTCAATAAGATGCAA-----CCTGGAATCCTACGGCAGTCCGCAGCAGCC	2422
QY	406	TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys	425
DB	2423	TACCTCGTAGCTTCGTGTCCCGTGGTGGCGCACATCTCTGGCGAGGTGTGAAGAGATGTC	2482
QY	426	LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys	445
DB	2483	TTTGACCTC-----TTAGGACCCCATCTGAACAACCTCCGTCCTTGACTATGAA	2530
QY	446	AlaPheCys-----AspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla	462
DB	2531	GCCACCTGTCAGGCCCTTGATCTTCGCCGCTACGGTCCTTCTACTCCACGGCACAGGCA	2590
QY	463	ValPheTyrThrPheValPheArgHisLysGlnLeu-----LeuSerGlyAsn	478
DB	2591	TTGCTGTATATCTTCTGCTCCGCTCCGCTGGCGAGACCTGACGACCGCAGCGATGGAGGCGAT	2650
QY	479	LeuLysGluGlyLeuGlnTyrLeu-----GlnSerLeuAsnPhe-----	491
DB	2651	ACCCCTGAACAAATTGATGAACCTGGAGCCAGAGACATCACGTTCCCTCCCTCGGTCAAG	2710
QY	492	-----GluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer	508
DB	2711	GAGGTCTCTACCAAGCAATCCATTCCAAACTGAACCCCTCAAAGTATGCTCGCGCGG	2770
QY	509	ValValAsnPhePheAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle	528
DB	2771	ATTGTGTGAATTTCGCCCGTATGGCTCAGCACCTCAACTTCATGTACGTGTTTCAGCATT	2830
QY	529	IleGluArgAsnAsnArgGlnMetLeuProValIleArgSer-----	542
DB	2831	CTCAGACGACAAACAAAGACTGCGGCTCTCCTCGTTCGGAGCATCTCAGCCATGGCAGAC	2890

QY	543		-----ThrAlaGlyAspSerValGlnIleCysThr	553
Db	2891	C CGCGCTT CAGCCATGTTC GACGAGAAACTCGTGCCGGCATGACCTGGTTAT-----		2944
QY	554	AsnProLeuAspThrPhePropheAspProCysValLeuLysArgSerLysLysPhe	573	
Db	2945	----CAGCTTGATGCCTATTTCCTTTGACC CATACCAGCTACCTCGCAGCCGCGTTGG	3001	
QY	574	IleAspProfileTyrglnValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLys	593	
Db	3002	CTGGAGGATGATTACGTCCTACTGGCGTGGTATCCCCTGGCGTGGACACCAAGGAC----	3055	
QY	594	LysProMetLysLysAspileValGluAspGluAspAspPheLeuLysGlyGluVal	613	
Db	3056	-----GAGGCGGATTCCGACAGTAGCACCTCGCGGAGCGA	3091	
QY	614	ProGlnAsnAspThrValIleGlyIleThrProSerSerPheAsp-ThrHisPheArgSe	633	
Db	3092	TCCGAAGGAGAC---CTGAGCGACGTCCTACTGAGACCGAGAAGACGACTGAATACTCTGA	3148	
QY	633	rProSerSerSerVal	638	
Db	3149	ACCAAACCTTGT	3164	
RESULT 8				
US-10-128-714-6207				
; Sequence 6207, Application US/10128714				
; Publication No. US20030119013A1				
; GENERAL INFORMATION:				
; APPLICANT: Jiang, Bo				
; APPLICANT: Hu, Wengqi				
; APPLICANT: Tishkoff, Daniel				
; APPLICANT: Zamudio, Carlos				
; APPLICANT: Eroshkin, Alexey M				
; APPLICANT: Lemieux, Sebastien M				
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus				
; TITLE OF INVENTION: Methods of Use				
; FILE REFERENCE: 10182-018-999				
; CURRENT APPLICATION NUMBER: US/10/128,714				
; CURRENT FILING DATE: 2002-04-23				
; PRIOR APPLICATION NUMBER: US 60/285,697				
; PRIOR FILING DATE: 2001-04-23				
; PRIOR APPLICATION NUMBER: US 60/287,066				
; PRIOR FILING DATE: 2001-04-27				
; PRIOR APPLICATION NUMBER: US 60/295,890				
; PRIOR FILING DATE: 2001-06-05				
; PRIOR APPLICATION NUMBER: US 60/303,899				
; PRIOR FILING DATE: 2001-07-09				
; PRIOR APPLICATION NUMBER: US 60/316,362				
; PRIOR FILING DATE: 2001-08-31				
; NUMBER OF SEQ ID NOS: 8603				
; SOFTWARE: PatentIn version 3.1				
; SEQ ID NO 6207				
; LENGTH: 2139				
; TYPE: DNA				
; ORGANISM: Aspergillus fumigatus				
US-10-128-714-6207				

QY 33 AsnMetArgAlaLeuGluAsnAspPhePheAsnSer-----ProProArgLysThrVal 50  
Db 175 CCCTCGTCTCGGTCGCTCCGATGGCAGCGATGGCAGCCCGCATCCAAAGGAAGCGGGTT 234  
QY 51 Arg-----PheGlyGlyThrValThr 57  
Db 235 CGGGTCAATTGTGATAAGGATGTGGAAATGAGAGAAAGGTCTTGGAGTGGTCAAGAGAAA 294  
QY 58 GluValLeuLeuLysTyrLysLysGly----- 66  
Db 295 GAAATGGCTGTGAGTACGGAAAGAGAGCGCCCGCGTGGTTCGTGAAGAGTGGCAGAGCT 354  
QY 67 -----GluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 78  
Db 355 ATCCAGCGGCATGTCTCTGGAACGGATAGTGAAGCCCTACGATCGTATCAAAAGAGATTTC 414  
QY 79 ---LeuAspProAspIleLysAspAspGln----- 87  
Db 415 TCGGCAGATCCTCGCCGCTGTGGACGAGATGGCTTGCTCCCGAAGATTACCAACGCAC 474  
QY 88 -----IleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAsp 105  
Db 475 ACTACATTGAAGCATCATCTCATGGGTCTTCTGTCCAAATGTTGGTCTGGTGGATCGCAAT 534  
QY 106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrVal 125  
Db 535 TGCAACGGCTTGGTCAATGAGTCCTGCACAGTGAATGGCTCGGGCGGATGAGTCGTAT 594  
QY 126 ValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145  
Db 595 ATCAAGTTGTACATTTCGATTTCTAGGTAACCTTGCGGCAGCCCCAGGGAAGCTATCTCGGG 654  
QY 146 ProCysLeuSerMetIleAlaSerHisPhe-----Val 156  
Db 655 GCGGTCTTGAAGATGTTGGCCAAACAATTTCGGGGAGATCCCAAAGGGGACGGCAAGCTA 714  
QY 157 ProProArgValIleIleLysGluGlyAspValAspValSerAspSerAspGluAsp 176  
Db 715 CCTGGCTACGCCCCAGTCCCAAGCTGCAGAGATC----- 747  
QY 177 AspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyr 196  
Db 748 -----TATAGCAGAACTCATATGGCTCTACGTCTCATGTGATGCAGTTG 789  
QY 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArg 216  
Db 790 ATTCTTCTGGTAGTGGGACCCTATCCGCCATCTTGTCTATGCAGTTTCCCTTCGATACG 849  
QY 217 LysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe 236  
Db 850 GATTCCGCAAGGCCAACATTGCGTACACGCGGAATCTCTTAAAGGTCATAGTTACGCC 909  
QY 237 ProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspVal 256  
Db 910 CCTGAGCTCCAAGCCGACATTCTAGCTCTCATCACGGAAAGTCGTGAAGATCGATGTC 969  
QY 257 AsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGly 276  
Db 970 CAGATT---CAGGTAGATATGGAAGACATCGAGGACGAAGTGGCGGAGATGTTCTCCAC 1026  
QY 277 ThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThr 296  
Db 1027 GCAGTCTCGCCTGAAGCCACGATGTTGGAGGATGACGAGAATGAAGTTGACAATGCATCC 1086  
QY 297 LysAlaGlyProGluArgLeuAsp-----GlnMetValHisProValAlaGlu--- 312  
Db 1087 GTCGTACGCATGATTCCTTAGACGCTGGAGTCCCGGAGTCAAGGGCCATCAAAAGACAAC 1146  
QY 313 -----ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330  
Db 1147 ATACTCAAATTGGATGGCATGATCGATTTCGTGTTTGAATACTACGCCCGCCCT---TTC 1203  
QY 331 ValAspGlyLysValAspAsnGlyLysThrLys---AspLeuTyrArgAspLeuIleAsn 349

Db 1204 ACCACGGGCACACTTTGGATGACAAGGAGAATGCACCTGGATTTA-----CTTCTGAGC 1254  
QY 350 IlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPhe 369  
Db 1255 CATTTCCAAAGTATTATCCTTCGACCTATCGCTCCGCCACTCTCAATTCTCTCTCTTC 1314  
QY 370 TyrLeuCysSerPheLysLeu-----GlyPheAlaGluAlaPheLeuGluHis 385  
Db 1315 CACITCTCTCCAGTCGTCGCCTATTCTGGTGGATCGATTGCGCAACCTGTGTTCAGCTC 1374  
QY 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn 405  
Db 1375 ATCTTCAATAAGATGCAA-----CCTGGAATCCTACGGCAGTCCGCAGCAGCC 1422  
QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425  
Db 1423 TACCTCGCTAGCTTCGTTGCCCGTGGTGGCGCACATCTCTGGCGAGGTGTGAAGATGTC 1482  
QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445  
Db 1483 TTTGACCTC-----TTAGGCACCCATCTGAACAACCTCCGTTCTTGACTATGAA 1530  
QY 446 AlaPheCys-----AspValAlaLeuHisGlyPropheTyrSerAlaCysGlnAla 462  
Db 1531 GCCACCTGTCGAGGCCCTGATCTTCGCGCTACGGTCTTCTACTCCACGGCACAGGCA 1590  
QY 463 ValPheTyrThrPheValPheArgHisLysGlnLeu-----LeuSerGlyAsn 478  
Db 1591 TTGCTGTATATCTTCTGCTTCGCTCGCGAGACCTTGACGACCGCAGCGATGGAGGCGAT 1650  
QY 479 LeuLysGluGlyLeuGlnTyrLeu-----GlnSerLeuAsnPhe----- 491  
Db 1651 ACCCCTGAACAATAATTGATGAACCTGGAGCCAGAAAGACATCACGTTCCCTCCCTCGGTCAAG 1710  
QY 492 -----GluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508  
Db 1711 GAGGTCCTTCACCAAGCAATCCATTCCAAACTGAACCCCTCAAAGTATGCTCGCCGCG 1770  
QY 509 ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528  
Db 1771 ATTGTGTGAATTTCGCCCGTATGGCTCGCGTCTCTCTCGTTCCGGAGCATCTCAGCCATTCAGCAT 1830  
QY 529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSer----- 542  
Db 1831 CTCGAGAGCAACAAAGACTGCGCGTCTCTCTCGTTCCGGAGCATCTCAGCCATGGCAGAC 1890  
QY 543 -----ThrAlaGlyGlyAspSerValGlnIleCysThr 553  
Db 1891 CCGCGCTTCAGCCATGTTCGAGCGAGAAAACCTCGTCCGCGCATGACCTGGGTAT----- 1944  
QY 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPhe 573  
Db 1945 ---CAGCTTGATGCCTATTTCCTTTGACCCATACCAGTACCTCGCAGCCGCGTTGG 2001  
QY 574 IleAspProIleTyrGlnValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLys 593  
Db 2002 CTGGAGGATGATTACGTCCACTGGCGTGGTATCCCTGGCGTGGACGACGAGGAC----- 2055  
QY 594 LysProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluVal 613  
Db 2056 -----GAGGCGGATTTCGGACAGTGACCTTCGACGGGAGCGAA 2091  
QY 614 ProGlnAsnAsp 617  
Db 2092 TCCGAAGGAGAC 2103

RESULT 9  
US-10-128-714-7207  
; Sequence 7207, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo



; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7207
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-7207

Alignment Scores:
Pred. No.: 7,48e-59 Length: 2139
Score: 563.00 Matches: 177
Percent Similarity: 44.03% Conservative: 133
Best Local Similarity: 25.14% Mismatches: 254
Query Match: 16.54% Indels: 140
DB: 15 Gaps: 23

US-09-932-678-2 (1-651) x US-10-128-714-7207 (1-2139)

QY 13 AspAlaAlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSer 32
Db 115 AACAGCGCGGAGCTGCCGGGCTAAACACAAGTTGGCCGACATGGATCTTTCGCTGTCA 174
QY 33 AsnMetArgAlaLeuGluAsnAspPhePheAsnSer-----ProProArgLysThrVal 50
Db 175 CCCTCGTCCTCGGTGCTTCGCGATGGCAGCGGATGGCACGCCCACTCCAAGGAAGCGGTT 234
QY 51 Arg-----PheGlyGlyThrValThr 57
Db 235 CGGGTTCAATTGATAAGGATGTGAAATGAGAGAAGCGTCTTGGAGTGGTCAAGAGAAA 294
QY 58 GluValLeuLeuLysTyrLysLysGly-----GluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 78
Db 295 GAAATGGCTGTGAGTACGGAAAAGAGCGGCCGCGTGGTTCGTGAAAGTGGCGCAGAGCT 354
QY 67 -----GluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 78
Db 355 ATCCAGCGGCATGTCTCTGGAACGGATAGTGAAGCCTACGATCGTATCAAAGAGATTTC 414
QY 79 ---LeuAspProAspIleLysAspAspGln-----87
Db 415 TCGGCAGATCCTCGCGCTGGACGACGAAGATGGCTTGCTCCCGAAGATTACCAACGCAC 474
QY 88 -----IleLeuAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAsp 105
Db 475 ACTACATTGAAGCATCATCTCATGGGTCTTCTGTCCAATGTTGCGTCCGTGGATCGCAAT 534
QY 106 PheGluGlnLeuIleSerIleLeuArgLeuProTrpLeuAsnArgSerGlnThrVal 125
Db 535 TGCAACGGCTTGGTCAATGCAGTCTCTGCACAGTGAATGGCTCGGGCGGATGAGTCGTAT 594
QY 126 ValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145
Db 595 ATCAAGTGTACATTCCGATTTCTAGGTAACTTGGGAGACCTTGGGAGAGCTATCTCGGG 654

QY 146 ProCysLeuSerMetIleAlaSerHisPhe-----Val 156
Db 655 GCGGTCTTGAAGATGTTGGCCAACAATTTCCGGGAGATCCCAAAGGGACCGGGCAAGCTA 714
QY 157 ProProArgValIleLeuLysGluGlyAspValAspValSerAspSerAspGluAsp 176
Db 715 CCTGGCTACGCCCCAGTCCAAGTGCAGAGATC-----747
QY 177 AspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyr 196
Db 748 -----TATAGCAGAACTCATATGGCTCTACGTCATGTGATGAGTTG 789
QY 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArg 216
Db 790 ATTCCTTCTGGTAGTGGACCCCTATCCCCCATCTTGTCTATGCAGTTTCCCTTCGATACG 849
QY 217 LysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe 236
Db 850 GATTCCGCAAAGCCCAACATTGCGTACACGCGGAATCTTCTAAAGGTCATTAGTTACGCC 909
QY 237 ProThrLeuArgHisGluIleLeuGluLeuIleLeuLysLeuLeuLysLeuVal 256
Db 910 CCTGAGCTCCAAGCCGACATTTAGCTCTCATCAGGAGAAAGTCGTGAAGATCGATGTC 969
QY 257 AsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGly 276
Db 970 CAGATT--CAGGTAGATATGGAAGACATCGAGGACGAAGTGGCGGAGGATGTTCTCCAC 1026
QY 277 ThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThr 296
Db 1027 GCAGTCTCGCTGAAGCCACGATGTTGGAGATGACGAGATGAAGTTGACAATGCATCC 1086
QY 297 LysAlaGlyProGluArgLeuAsp-----GlnMetValHisProValAlaGlu--- 312
Db 1087 GTCGTAGCGATGATTCCTTAGACGCTGGAGTCCCGGCGAGTCAAGGCCATCAAAGACAAC 1146
QY 313 -----ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330
Db 1147 ATACTCAAATGGATGGCATGATCGATTCCGCTGTTTGAATACTACGCCCCCGCCT---TTC 1203
QY 331 ValAspGlyLysValAspAsnGlyLysThrLys---AspLeuTyrArgAspLeuIleAsn 349
Db 1204 ACCACGGGCACTTTGGATGACAAGGAGAATGCACCTGGATTAT-----CTTCTGAGC 1254
QY 350 IlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPhe 369
Db 1255 CATTTCCAAAGTATTATCCTTCCGACCTATCGTCCCGCCACTCTCAATTCTCTCTCTTC 1314
QY 370 TyrLeuCysSerPheLysLeu-----GlyPheAlaGluAlaPheLeuGluHis 385
Db 1315 CACTTCTCCCAGTCGTCGCCTATTCTGGTGGATCGATTGCGCGCAACCTGTGTTCAGCTC 1374
QY 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn 405
Db 1375 ATCTTCAATAAGATGCAA-----CCTGGAATCTACGGCAGTCCGACGACGCC 1422
QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
Db 1423 TACCTCGCTAGCTTCGTGCCCCGTGGTGGCGCACATCTCTGGCGAGGTGTAAGAGATGTC 1482
QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445
Db 1483 TTTGACCTC-----TTAGGCACCCATCTGAACAACCTCCGCTTTGACTATGAA 1530
QY 446 AlaPheCys-----AspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462
Db 1531 GCCACTGTGAGGCGCCTGATCTTCGCCGCTATCGCGCTTCTTCTACTCCACGGCACAGCA 1590
QY 463 ValPheTyrThrPheValPheArgHisLysGlnLeu-----LeuSerGlyAsn 478
Db 1591 TTGCTGTATATCTTCTGTCTTCCGCTTCCGCTGGCAGACCTGACGACCGCAGCGATGGAGGCGCAT 1650

QY 479 LeuLysGluGlyLeuGlnTyrLeu-----GlnSerLeuAsnPhe----- 491  
Db 1651 ACCCCTGAACAAATTGATGAACCTGGAGCCAGAGACATCACGTTCCCTCCCTCGGTCAAG 1710  
QY 492 -----GluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508  
Db 1711 GAGGTCCTTCACCAAGCAATCCATTCCAAACTGAACCCCTCAAAGTATGTCGCCGCG 1770  
QY 509 ValValAsnPheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528  
Db 1771 ATTGTGTCTGAATTCGCCCGTATGGCTCAGCACCTCAAACTTCATGTACGTGTTTCAGCATT 1830  
QY 529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSer----- 542  
Db 1831 CTCGAGACGAACAAAAGACTGCGCGTCTCCTCGTTCGGGAGCATCTCAGCCATGGCAGAC 1890  
QY 543 -----ThrAlaGlyGlyAspSerValGlnIleCysThr 553  
Db 1891 CCGCGCTTCAGCCATGTCGAGCGAGAAACTCGTGCCGCGCATGACCTGGGTTAT----- 1944  
QY 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPhe 573  
Db 1945 ---CAGCTTGATGCCTATTTCCTTTGACCCATACCAGCTACCTCGCAGCCCGCGTGG 2001  
QY 574 IleAspProIleTyrGlnValTrpGluAspMetSerAlaGluLeuGlnGluPheLys 593  
Db 2002 CTGGAGGATGATTACGTCCACTGGCGTGTGTATCCTCGGTGGACGACGAGAC----- 2055  
QY 594 LysProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluVal 613  
Db 2056 -----GAGGCGGATTTCGGACAGTGACCTCGACGGGAGCGAA 2091  
QY 614 ProGlnAsnAsp 617  
Db 2092 TCCGAAGGAGAC 2103  
RESULT 10  
US-10-128-714-1207  
; Sequence 1207, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1207  
; LENGTH: 1878  
; TYPE: DNA  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-1207  
Alignment Scores: 9.14e-59 Length: 1878  
Pred. No.: 561.50 Matches: 174  
Score:

Percent Similarity: 45.71% Conservative: 124  
Best Local Similarity: 26.69% Mismatches: 253  
Query Match: 16.50% Indels: 101  
DB: 15 Gaps: 21  
US-09-932-678-2 (1-651) x US-10-128-714-1207 (1-1878)  
QY 26 AlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsnSerPro 45  
Db 10 GCGTCTGGAGTGTGTCAGAGAAAGAAATGGCTGTGAGTACGGAAAGAGCGCGCGGTG 69  
QY 46 ProArgLysThrValArg-----PheGlyGlyThrValThrGluVal 59  
Db 70 GTTCGTGAAGAAGTGGCAGAGCTATCCAGCGGCATGTCTCTGGAACGGATAGTGAAGCC 129  
QY 60 LeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeu 79  
Db 130 TACGATCGTATCAAAGAGATTTTCTCGGCAGATCTCTCGCCGTCTCGACGAAGATGGCTTG 189  
QY 80 AspPro---AspIleLysAspAspGlnIleIle---AsnTrpLeuLeuGluPheArgSer 97  
Db 190 CCTCCCGAAGATTACCAACGACACTACATTTGAAGCATCATCTCATGGGTCTTCTGTCC 249  
QY 98 SerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleLeuArgLeuPro 117  
Db 250 AATGTTGCGTCGCTGGATCGCAATTGCAACGGCTTGTCAATGCAGTCTGCACAGTGAA 309  
QY 118 TrpLeuAsnArgSerGlnThrValValGluGluGlyLeuAlaPheLeuGlyAsnLeuVal 137  
Db 310 TGGCTCGGCGGGATGAGTCGTATATCAAGTTGTACATTCGATTTCTAGGTAACTTGGC 369  
QY 138 SerAlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPhe----- 155  
Db 370 GCAGCCGAGGGAAGCTATCTCGGGGGTCTTGAAGATGTTGGCCACAATTTTCGGGGAG 429  
QY 156 -----ValProProArgValIleIleLysGluGlyAspValAsp 168  
Db 430 ATCCCAAAGGGGACGGGCAAGCTACCTGGCTACGCCCCAGTCCAAGCTGCAGAGATC--- 486  
QY 169 ValSerAspSerAspAspGluAspAsnLeuProAlaAsnPheAspThrCysHisArg 188  
Db 487 -----TATAGCAGAACTCATATG 504  
QY 189 AlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeu 208  
Db 505 GCTCTACGTCATGTGATGCAGTTGATTCCTTCTGTTAGTGGGACCCCTATFCCCCCATCTTG 564  
QY 209 ValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsn 228  
Db 565 TCTATGCAGTTTCCCTTCGATACGGATTCCGCAAGGCCAACCAATTCGCTACACGCGGAAT 624  
QY 229 LeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIle 248  
Db 625 CTTCCTAAAGGTCATTAGTTAGCCCTCGAGCTCCAAGCCGACATTTCTAGCTCTCATCAG 684  
QY 249 GluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGlu 268  
Db 685 GAGAAAGTCGTGAAGATCGATGTCCAGATT---CAGGTAGATATGGAAGACATCGAGGAC 741  
QY 269 ThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGlu 288  
Db 742 GAAGTGGCGGAGGATGTTCTCCACGCGAGTCTCGCTGAAGCCACGATGTTGGAGGATGAC 801  
QY 289 AspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAsp-----Gln 305  
Db 802 GAGAAATGAAGTTGACAATGCATCCGTCGTAGCGATGATTCCCTAGACGTGGAGTCCCGG 861  
QY 306 MetValHisProValAlaGlu-----ArgLeuAspIleLeuMetSerLeuValLeu 322  
Db 862 CGAGTCAAGGCCCATCAAAGACAACATACTCAAATGGATGGCATGATCGATTGCTGTTT 921  
QY 323 SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLys--- 341

Db 922 GAATACTACGCCCGCCT---TTCACCACGGGCACTTTGGATGACAAAGGAGAAATGCAC TG 978

QY 342 AspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSer 361

Db 979 GATTTA-----CTTCTGAGCCATTTCCAAAGTATTATCCTCTCCGACCTATCGCTCC 1029

QY 362 CysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeu-----Gly 377

Db 1030 CGCCACTCTCAATTCCTCTTCCACTTCTCCAGTCGTGCGCTATTCTGTTGGATCGA 1089

QY 378 PheAlaGluAlaPheLeuGluHisLeuTyrLysLeuGlnAspProSerAsnProAla 397

Db 1090 TTTGCCGCAACCTGTGTTTCAGCTCATCTTCAATAAGATGCAA-----CCTGGA 1137

QY 398 IleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIle 417

Db 1138 ATCCTACGGCAGTCCGACGAGCTACCTCGCTAGCTTCGTTGCCGTTGGTGGCACATC 1197

QY 418 ProLeuIleThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeu 437

Db 1198 TCTGGCGAGGTTGTAAGAGATGTCTTTGACCTC-----TTAGGCACCCATCTG 1245

QY 438 AsnAsnGlnAspSerGlyThrLysAlaPheCys-----AspValAlaLeuHisGly 454

Db 1246 AACAACTCCGTCTTGACTATGAAAGCCACCTGTGAGGCCCTGATCTTCGCGCTACGGT 1305

QY 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474

Db 1306 CCTTTCTACTCCACGGCACAGGCATTGCTGTATATCTTCTGCTTCGCTGGCGAGACCTG 1365

QY 475 -----LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeu-----GlnSer 488

Db 1366 ACGACCGCAGCGATGGAGGGCGATACCCCTGAACAAATTGATGAACCTGGAGCCAGAAGAC 1425

QY 489 LeuAsnPhe-----GluArgIleValMetSerGlnLeuAsn 500

Db 1426 ATCACGTTCCCTCCCTCGGTCGAGGAGGTCTCTCACCAAGCAATCCATTCCAACTGAAC 1485

QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520

Db 1486 CCCCTCAAAGTATGCTCGCGCGGATTGTGTCTGAATTCGCCGTTATGGTCTCAGCACCTC 1545

QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540

Db 1546 AACTTCATGTACGTGTTTCAGCATTTCTCGAGACGAACAAAGACTGCGCGTCTCCTCGTTC 1605

QY 541 ArgSer-----ThrAlaGly 545

Db 1606 CGGAGCATCTCAGCCATGCGCAGACCCGCGCTTCAGCCATGTCGAGCGAGAAACTCGTGCC 1665

QY 546 GlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePropheAspProCys 565

Db 1666 GCGATGACCTGGGTTAT-----CAGCTTGATGCTATTTTCCCTTGACCCATAC 1716

QY 566 ValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspMetSer 585

Db 1717 CAGCTACCTCGCAGCCGCGTTGGCTGGAGGATGATTACGTCCACTGCGGTGGTATCCCT 1776

QY 586 AlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAsp 605

Db 1777 GGCGTGGACGACGAGGAC-----GAGGGCGATTCTG 1806

QY 606 AspAspPheLeuLysGlyGluValProGlnAsnAsp 617

Db 1807 GACAGTGACCTCGACGGGAGCGAATCCGAAGGAGAC 1842

RESULT 11

US-10-128-714-2207  
; Sequence 2207, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi

; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2207  
; LENGTH: 1830  
; TYPE: DNA  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-2207

Alignment Scores:  
Pred. No.: 2,07e-56 Length: 1830  
Score: 542.50 Matches: 169  
Percent Similarity: 44.79% Conservative: 123  
Best Local Similarity: 25.92% Mismatches: 243  
Query Match: 15.94% Indels: 117  
DB: 15 Gaps: 21

US-09-932-678-2 (1-651) x US-10-128-714-2207 (1-1830)

QY 26 AlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsnSerPro 45

Db 10 GCGTCTGGAGTGGTCAAGAGAAAGAAATGGCTGTGAGTACGGAAAGAGCGCCCGTG 69

QY 46 ProArgLysThrValArg-----PheGlyGlyThrValThrGluVal 59

Db 70 GTTCGTGAAGAAGTGGCAGAGCTATCCAGCGGCAATGCTCTGGAACGATAGTGAAGCC 129

QY 60 LeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeu 79

Db 130 TACGATCGTATCAAGAGATTTTCTCGGCAGATCCTCGCCGTCCTGGACGAAGATGGCTTG 189

QY 80 AspPro---AspIleLysAspAspGlnIleIle---AsnTrpLeuLeuGluPheArgSer 97

Db 190 CCTCCCGAAGATTACCAACGCACATACATTGAAGCATCATCTCATGGGTCTTCTGTGTC 249

QY 98 SerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPro 117

Db 250 AATGTTGCGTCGCTGGATCGCAATTGCAACGGCTTGGTCAATGCACTCCTGCACAGTGAA 309

QY 118 TrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuVal 137

Db 310 TGGCTCGGGCGGGATGAGTCGTATATCAAGTTGTACATTTCGATTTCAGGTAACCTTGC 369

QY 138 SerAlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPhe----- 155

Db 370 GCAGCCAGGGAAGCTATCTCGGGGCGGTCTTGAAGATGTTGGCCAACAATTTTCGGGAG 429

QY 156 -----ValProProArgValIleIleLysGluGlyAspValAsp 168

Db 430 ATCCCAAAGGGACGGGCAAGCTACCTGGCTACGCCCCAGTCCCAAGCTGCAGAGATC--- 486

QY 169 ValSerAspSerAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArg 188

Db 487 -----TATAGCAGAACTCATATG 504



Db	458	CGAGAGAAATTACGATGAGCTCGTCGGTTATTGTCACCCGAGAAGCATCTTAATCCTGAT	517
QY	83	IleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerIleMetTyrLeu	102
Db	518	---GAGGTTGCTATACTGTGACGACTTGAAGCGCGTTTCTGGAGCAGTGTCTCTATATA	574
QY	103	---ThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArg	121
Db	575	GATTCTGTTTCATCATGAATCCCTGCTCTTTGCTGTTCGAGAAATGAGCCTCTGGAAATTAT	634
QY	122	SerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThr	141
Db	635	GGAATTGATATTATGATGCACTGCTGGAACCTCATATCTTTGGCTGCTTCCCAATGGA	694
QY	142	ValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgValIle	161
Db	695	AAATACATCGATTGGTGTGTTGGAGATGCTTGTGAAATAATTTTGCCGCTTTTACCTC	754
QY	162	IleLysGluGlyAspValSerAspSerAspAspGluAspAspAsnLeuProAla	181
Db	755	TTA-----GATTCTGTGAGGCAGGAGAACGGAATTGAAAGA	790
QY	182	Asn-----PheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValPro	198
Db	791	AAAAATAAAGTTCTGTACGGGTACATGCAGCATTTGAAAGAAATAGCTGATTGGTGCCT	850
QY	199	SerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSer	218
Db	851	CTTGCAACCTTGCAGACTATCTCAATAGTTATCCAGAAATGCGAGTGTCTCAGTAAG	910
QY	219	GluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIle-----SerValTyrPhe	236
Db	911	GAAACTGAGATTGTCACTGTATGTCGAGAAATATGTTGAGGCTGGAGAGTGGTGCACTTGA	970
QY	237	ProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspVal	256
Db	971	GAAACTGTTGGTCTTACCATGCTGCTGCTGCTGTTGGATAGGTTGCTAGAGTTGGATGTG	1030
QY	257	AsnAlaSerArgGlnGlyIle-----GluAspAlaGluGluThrAlaThrGlnThrCys	274
Db	1031	GAAATTGGATGGGATGGAATCCTGCAAGAAGATGCT-----	1066
QY	275	GlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHis	294
Db	1067	-----AAAGGCATATTGAAATGGAGTTAGAGTTGCTAGAGTTGTAATTT	1108
QY	295	GluThrLysAlaGlyProGluArgLeuAspGlnMetVal-----	307
Db	1109	GCAGATGAG-----GATGAGAATTGTGATAGTATGCTCCGTCAGAGTTATTAATAGA	1162
QY	308	-----HisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu	322
Db	1163	AAAAATTGCAAGGCAACTTGGTTGTGAGAAATTAGATAGCCTAATG-----GTGCTG	1216
QY	323	SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp	342
Db	1217	GCATTTTGT-----CACCTGAATCCTGTCAAAGCAGTGGCGCTTGTACTGAG	1264
QY	343	LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCys	362
Db	1265	GTATTGCTACTCTACTAGCTTCATTTCAGAGAACCGTATTGAATGCGTACAAGTCAAAA	1324
QY	363	HisValGlnPhePheMetPheTyrLeuCysSer-----	373
Db	1325	TTTACCAGTATGTGATGTTCTATGCAATGTGCACCTGACCCCTGAAGGATGTGGTGAAA	1384
QY	374	PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro	393
Db	1385	TTTGGCATGGTCTTGCTGATATGTT-----GGATGTGAT	1420
QY	394	SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg	413
Db	1421	GTTAATCCCCCTATTACTAGGATGAGTGCTGCTGTATCTTGTAGTATTATTGACTCGT	1480
QY	414	AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu	433
Db	1481	GCAAAGTTCTTTCAGCTGCATTGGTCACTAACATCATACAAAAGCTTGGTTGATTGGTGT	1540
QY	434	HisIleTyr-----LeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAla	451
Db	1541	TATGCATATTGCAAGTTACGTGATTTTGTATATGAACCCACGAGCA-----	1585
QY	452	LeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHis	471
Db	1586	---CATCAAGTTTATTCTGGTGCCAGGCCAATATGATATATTCTGTGTTTTCGCATG	1642
QY	472	LysGlnLeuLeuSer---GlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsn	490
Db	1643	AGATCTTTGATGGATATTCTTAGGCTTAAA-----TTGCAGCTGCTTAAT	1687
QY	491	-----PheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer	508
Db	1688	ATGCCCATGGAGCAATCTTGAAACATAAATGAGTCCCTTGAAGGTGTGTTGCCTACT	1747
QY	509	ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle	528
Db	1748	GTAGTAGTGAATTTCTTAGACAGGCAAGGCTGCTCACTTTTCATGGCGTCAGAGTCA	1807
QY	529	IleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSer	548
Db	1808	TTTGTTTTGTGATATGCTTGTGAGTCTGATCTTCCAGGCTTTTGGCGGATGATAGA	1867
QY	549	ValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLeuLys	568
Db	1868	-----CTTGACATGTTTTCCTCAATTTGATCCATGTTTGTGTTGAAG	1906
QY	569	ArgSerLysLysPheIleAspProIleTyrGlnValTrp-----	581
Db	1907	AAAAGTGAAAGCTACATAAGGCCACATTTTGTTCGCTGGTCAAAGTCAGAACTACATAT	1966
QY	582	GluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleVal	601
Db	1967	GATGATGATGACAATGATGAACACTCAGTGATAGTGGCAGT-----	2005
QY	602	GluAspGluAspAspPheLeu-----LysGlyGluValProGlnAsnAspThr	618
Db	2006	GAATTATCAGATGATGATTTTATCGACAGGAACACAAAGGATATGATTGATGATATG	2065
QY	619	ValIleGlyIleThrProSerSerPheAsp	628
Db	2066	ATGGTGAGCGTTGAAGGTTTGTGATTTTAAC	2095
RESULT 13			
US-10-032-585-6770			
; Sequence 6770, Application US/10032585			
; Publication No. US20030180953A1			
; GENERAL INFORMATION:			
; APPLICANT: Terry, Roemer D.			
; APPLICANT: Bo, Jiang			
; APPLICANT: Charles, Boone			
; APPLICANT: Howard, Bussey			
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery			
; FILE REFERENCE: 10182-005-999			
; CURRENT APPLICATION NUMBER: US/10/032,585			
; CURRENT FILING DATE: 2001-12-20			
; NUMBER OF SEQ ID NOS: 8000			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 6770			
; LENGTH: 1671			
; TYPE: DNA			
; ORGANISM: Candida albicans			
US-10-032-585-6770			
Alignment Scores:			
Pred. No.:	4.19e-51	Length:	1671



Score:	499.50	Matches:	156
Percent Similarity:	46.61%	Conservative:	98
Best Local Similarity:	28.62%	Mismatches:	219
Query Match:	14.68%	Indels:	72
DB:	15	Gaps:	15
US-09-932-678-2 (1-651) x US-10-032-585-6770 (1-1671)			
QY	81	ProAspIleLysAspGlnIleAsnTrpLeuGluPheArgSerSerIleMet	100
Db	192	CCCCGAATCAATCACACTCCCCCAATTCAGATCCTTGAA---CGGACTCGTGCAAA	248
QY	101	TyrLeuThr-----LysAspPheGluGln-LeuIleSerIleIleLeuArgLeuPr	117
Db	249	TACCTCACGCTTGGACAACACCGCATGCAGCAACCTCATCTTTGCCATTTTGAGATACAA	308
QY	117	oTrpLeuAsnArgSerGlnThr-----ValValGluGluTyrLeuAlaph	132
Db	309	CTGGCTAGACATAGCCGATACCGATGCGGTACCACCGTTTCGCTGAATCGTATTCCCAATT	368
QY	132	eLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMetIleAl	152
Db	369	CCTCGTGTGATTGGTGTGCGCGCTCCCAAAATACCTCCATGAAGTGACAAACAACTTAT	428
QY	152	aSerHisPheValProArgValIleIleLysGluGlyAspValaspValSerAspSe	172
Db	429	CCAGGAATTGACTCGCCA-----	447
QY	172	rAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIl	192
Db	448	-----AATACGACCCACCACCACCGAAATACTCACCAA	479
QY	192	eIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysPh	212
Db	480	AATCATCAATAACGTTCCACATCAATCAACTCAATCCCACTCTTGCAAAACAACTT	539
QY	212	ePro---PheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuAr	231
Db	540	CCCCCACCATTTGCTCTCGACAAGGAACCTTACCAACTACGTGCAGAACCTAGTCCG	599
QY	231	gIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLe	251
Db	600	CGTCTTGGACTACACTCCAGAGTTACAATTAGTATATGGCAACTAATCATCGAGTCGAG	659
QY	251	uLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaTh	271
Db	660	TATAAAACTAGACGTTGAATCCAGAACGAA---CTTGACGACCTAGACGAC-----	708
QY	271	rGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGl	291
Db	709	-----GAAGAGATCGAAGATTTGATTAGTGGCGAAGAAGACTCCGA	749
QY	291	uThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAl	311
Db	750	CAGCGAAGACGATGGGAGGTACCACCCTCCATCAAAATCCCTCCTG-----	798
QY	311	aGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrVa	331
Db	799	-TCCAAACTCGATAGCATATTACGACACCTCTTGACTACCACCGCAACTCATTCACCAT	857
QY	331	lAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePh	351
Db	858	CGAA---GAACATAAACACCGGAACCGGTGTCAACCTCTTCAACACCATCAACTCCCTCTT	914
QY	351	eAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLe	371
Db	915	CAAAAGCCACATCTTGCCCAACCCACTTTACCAAACTGATCCAGTTATCTTATTCACAT	974
QY	371	uCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGl	391
Db	975	CACCAGTTCCAACTGAATTGGCCGACTCATACTTGGTATTGCTCATCGATGGCTTT	1034
QY	391	nAspProSerAsnProAlaIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLe	411

Db	1035	CAACCCAGCGGAAATCTTGGAAAAGCGTCTCAAGGCAATGCAGTACTTGTCTCATACAT	1094
QY	411	uAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAs	431
Db	1095	CGCCAGAGCCAAAACCTTTCCCGCCACCAGGTCATCTTCATTGTTCAGTTATCTTGTGG	1154
QY	431	nTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAl	451
Db	1155	TTGGCTCAACAAGTTTATCATCGAAAAGAGAACAGAGGTGAGTCTAGTGGCGACAAACA	1214
QY	451	aLeuHisGlyPro-----PheTyrSerAlaCysGlnAlaValPheTyrTh	466
Db	1215	AGTCGGGGGTATGGAACGATTCAAACTATTCTACCGCGCATTTTCAGGCATTGCTCTACAT	1274
QY	466	rPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLe	486
Db	1275	CTTCTGTTTCCGCCAC--GAACCTACTCTCCACCACCGAAACGACTGGGAGTGCACCT	1331
QY	486	uGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLe	506
Db	1332	CGACAAGTTC---TTCCAAAGAGCTATCCTTACCAAGTTTAAACCCGCTAAAGTTCTGTGA	1388
QY	506	uProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTy	526
Db	1389	CGAGACCGTGGTGTACATATTGGCCAAGTTGGCCACCAAACTCAATGTCTGTACTGCTA	1448
QY	526	rThrIleIleGluArgAsnAsnArgGlnMetLeu-----ProValIleArgSerTh	543
Db	1449	CTCCATCATCGAACACAATAAACCGCGAGAGAAATGTGCACACAAACGCATTGCCTTCGC	1508
QY	543	rAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePropheAs	563
Db	1509	AGTTGGCAACTTTAAACAGAAAACAAGAGTCTTGGACTTGAAGCCTACTTCCCATTCGA	1568
QY	563	pProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTrp---Gl	582
Db	1569	TCCATTGGTCTTGCCCATCAGCAAAATCCATCGTCGTCCTCAAACTACATAGAATGGTCAG	1628
QY	582	uAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGl	602
Db	1629	AGATACCCATGCCGAAGAA-----ACCGA	1652
QY	602	uAspGluAspAsp	606
Db	1653	TAGTGAAGACGAT	1665

RESULT 14  
US-09-938-842A-2204  
; Sequence 2204, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SCRIPl300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 2204  
; LENGTH: 1815  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-2204

Alignment Scores:				
Pred. No.:	1.15e-45	Length:	1815	
Score:	456.50	Matches:	168	
Percent Similarity:	42.56%	Conservative:	98	
Best Local Similarity:	26.88%	Mismatches:	257	
Query Match:	13.41%	Indels:	103	
DB:	9	Gaps:	23	
US-09-932-678-2 (1-651) x US-09-938-842A-2204 (1-1815)				
QY	42	PheAsnSerProProArgLysThrValArgPheGlyGlyThr-----Val 56		
Db	31	TTCAACTCAAAATGCAATGGACAACGCTGAATATATACCGACACAGATTGGTGTTCGCCGTT 90		
QY	57	ThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLysAsn 76		
Db	91	AGAAAGGCCCTTGCATCTGTCCAAATGGAGATACCGACGATTATAGTCAGCTTAAACA 150		
QY	77	GlnLeu---LeuAspProAspIleLysAspAsp-----GlnIleIleAsnTrpLeu 92		
Db	151	GTGATGTGCCTTACAGAAGACGCTGATTTTCGATGCAGTGGCACAGCTTGAGACAGTCTTG 210		
QY	93	LeuGluPheArgSerIleMetTyrLeuThr-----LysAspPheGluGln 108		
Db	211	AAAAGTCTATCAGTTTCCGTTGCTGGATAGATTAGTTTCATCATATAAGATCTCTTGAA 270		
QY	109	LeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGln---ThrValValGlu 127		
Db	271	GCTATGAGCTTG-----TGGTATCATAGTCACAGACCTAGTGTAAATGGAT 315		
QY	128	GluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCys 147		
Db	316	GCATTGGTGGACCTAATCAATCACTGGCCGTACTAGTGGAATAATCTGGATCCTTGT 375		
QY	148	LeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVal 167		
Db	376	TTGAATATGCTCGTAAGAAATTCAGTCAACCTACTTTTAAAGCATAAAGTTTCACAAACC 435		
QY	168	AspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHis 187		
Db	436	CAATTAGTCAAAAAGATGCGAGAA-----GTTTCATCCACGGGTGCAT 477		
QY	188	ArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIle 207		
Db	478	GCAGCCCTTCACAAGATTCTTATTAAATCTCTTGTGCTCCCTGGAAATTAGTGTCTATA 537		
QY	208	LeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHis 227		
Db	538	CTTGCGCAGAACATGCGTAAATTTGACAAAGGACCCCTCTATAGTGACATATGTGGAT 597		
QY	228	AsnLeuLeuArgIle-----SerValTyrPheProThrLeuArgHisGluIleLeuGlu 245		
Db	598	AACCTGTGTGAGTTGGAGAAATAGCTCAATCGGAGAAAGTTGTTGGCAGCGTGATTCTTATG 657		
QY	246	LeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAsp 265		
Db	658	ATGGTAATGGAGAGGATGCTAGATTTAGATTG-----690		
QY	266	AlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285		
Db	691	-----GTAAGTGGATGTGATGACTCTAATGGAGGCACTGAATGAAGGAGACGAGTTTCCAGTGGGG 789		
QY	286	MetAspGluAspGluThrGluHisGluThrLysAlaGlyProGlu-----301		
Db	730	ATGGAACTTGAAGATGCAGTTGAAAGCACTATGAATGAAGGAGACGAGTTTCCAGTGGGG 789		
QY	302	ArgLeuAspGlnMetVal-----HisProValAlaGluArgLeuAspIleLeuMet 318		
Db	790	GCTCTAAAACAAATACTTCCAGTGGAAATGTAGTCTCTGAACCTGTGGACAAATGATG 849		
QY	319	SerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGly 338		

Db	850	GTCTTATTTTTTCATCATCTAGAAATCC---TGTCAAAAACTCTGATCGTTGGAT----- 900		
QY	339	LysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThr 358		
Db	901	-----GAGGTATTTGAAATCCTCTTTAAGTCGTCGAGAACTATATTTCTGAACACA 951		
QY	359	HisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLys----- 375		
Db	952	TACAAAACAAAATTTTCACAGTTTCTGATGTTCTATGCTATGCTACTAGATCCTCGAAAAAT 1011		
QY	376	-----LeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393		
Db	1012	TGTGGTGTGAGATTGGCCAGTAAAGCTGTTGGAC-----ATATATCTCTCC 1056		
QY	394	SerAsnProAla---IleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAla 412		
Db	1057	AGCAACACATGTCGACTTACTAGGATGAGTGCAGTGGCTTATCTAGCTAGCTATTGTCT 1116		
QY	413	ArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrp 432		
Db	1117	CGTGAAAAGTTTTCGCTGCTCTTTTGTGGCTAGCATGTTGAAAAAGATTGGTGGCAGAG 1176		
QY	433	LeuHisIleTyrLeuAsn-----AsnGlnAspSerGlyThrLysAlaPheCysAspVal 450		
Db	1177	TGTGCGGAATATTCGGAACATGCAATGATGATGTGAAGCCAGAACGA----- 1224		
QY	451	AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470		
Db	1225	-----CATCAAGTGTCTTATCTGGATGTGAGGCAATCTTGATGTGCTATGCTTCGGA 1278		
QY	471	HisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSer----- 488		
Db	1279	ATGAGATCCATAGTGGAG-----ATTCTCGCTTTCAATCGCAGTTT 1320		
QY	489	LeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508		
Db	1321	AGATCATTTGGAGTCAATTTATCTCACAAACTAAACCCACTACTGGTGTGTCTTCCATCT 1380		
QY	509	ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528		
Db	1381	GTAGTTTCGAGTTCCTTAA-ACAAGCCAAAGCTGGTGGTCTGTTCATTGTTCTCAGAATC 1439		
QY	529	IleGlu-ArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspse 548		
Db	1440	CTTCATTTTCGATGACCTACACGAGTCTGAGCTCTCTCGTCTTTTGGTGGCTTTGAAAG 1499		
QY	548	rValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLeuLys 568		
Db	1500	G-----CTTGACACATTCTTCCCGTTTGACCCGCTGCTTGTGAA 1538		
QY	568	sArgSerLysLysPheIleAspProIleTyrGlnValTrp-----581		
Db	1539	AATGTCTAGCAGCTACATCTCCCCGAACTTCAACTTCTGGTCAATGGTGAAACGACTTA 1598		
QY	582	---GluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600		
Db	1599	TGGAGAAAGATGGTGACGAAGAGCTTTGTGATGAGGTCATAGTGAATGGAGATGCAGACAG 1658		
QY	600	eValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620		
Db	1659	TGCGAGGAACCTGACGATGACGTT-----GAACCTTGATAGTGAACACGAT 1709		
QY	620	eGlyIleThrPro 624		
Db	1710	GTCTACAACTCCG 1722		

RESULT 15  
US-09-938-842A-2204  
; Sequence 2204, Application US/09938842A  
; Publication No. US20040009476A9  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel

```
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2204
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2204
```

```
Alignment Scores:
Pred. No.:      1.15e-45      Length:      1815
Score:          456.50      Matches:      168
Percent Similarity: 42.56%      Conservative: 98
Best Local Similarity: 26.88%      Mismatches: 257
Query Match:      13.41%      Indels:      103
DB:              11          Gaps:       23
```

US-09-932-678-2 (1-651) x US-09-938-842A-2204 (1-1815)

QY	42	PheAsnSerProProArgLysThrValArgPheGlyGlyThr-----Val 56
Db	31	TTCAACTCAATGCAATGGACAACGCTGAATATACCGACACAGATTTGGTGTGTTGCCGTT 90
QY	57	ThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsn 76
Db	91	AGAAAGGCCCTTGCACTGTCCAAAATGGAGATACCGAGATTATAGTCAGCTTAAACA 150
QY	77	GlnLeu---LeuAspProAspIleLysAspAsp-----GlnIleLeuAsnTrpLeu 92
Db	151	GTGATGTGCCTTACAGAAGACGCTGATTTCGATGCAGTGGCACAGCTTGAGACAGTCTTG 210
QY	93	LeuGluPheArgSerSerIleMetTyrLeuThr-----LysAspPheGluGln 108
Db	211	AAAAGTCTATCAGTTTCCGTGTGCTTGGATAGATTAGTTTCATCATAAAGATCTTCTTGAA 270
QY	109	LeuIleSerIleLeuArgLeuProTrpLeuAsnArgSerGln--ThrValValGlu 127
Db	271	GCTATGAGCTTG-----TGTATCATAGTCACAGACCTAGTGTAAATGGAT 315
QY	128	GluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCys 147
Db	316	GCATTGGTGACCTAATCATATCACTGCGCGCTACTAGTGGGAAAATATCTGGATCCTTGT 375
QY	148	LeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVal 167
Db	376	TTGAATATGCTCGTAAGAAATTTCAGTCAACCTACTTTTAAGCATAAAGTTTCACAAACC 435
QY	168	AspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHis 187
Db	436	CAATTAGTCAAAAAGATGCAGAA-----GTTTCATCCACGGGTGCAT 477
QY	188	ArgAlaLeuGlnIleAlaArgTyrValProSerThrProTrpPheLeuMetProIle 207
Db	478	GCAGCCCTTCACAAGATTTCTATTAAATTCCTTGTCTCCCTGGAATTTAGTCTATA 537
QY	208	LeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHis 227
Db	538	CTTGGCAGAACATGCGTAAATTGACAAAAGGACCCCTTCTATAGTGACATATGTGGAT 597
QY	228	AsnLeuLeuArgIle-----SerValTyrPheProThrLeuArgHisGluIleLeuGlu 245
Db	598	AACGTGTGAGGTGGAGAAATAGCTCAATCGGAGAAGTTGTTGGCAGCGTGATCTTATG 657

QY	246	LeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAsp 265
Db	658	ATGGTAATGGACAGGATGCTAGATTTAGATTTG----- 690
QY	266	AlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285
Db	691	-----GTAAGTGGATGTGATGACTCTAATGGAGGCATGTTGAT 729
QY	286	MetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGlu----- 301
Db	730	ATGGAACCTTGAAGATGCAGTTGAAAGCACATATGAATGAAGAGACGAGTTTCCAGTGGGG 789
QY	302	ArgLeuAspGlnMetVal-----HisProValAlaGluArgLeuAspIleLeuMet 318
Db	790	GCTCTAAACACAAAATACTTCAGGTGGAATGTAGTCTCTGAACTGTTGGACAAATTGATG 849
QY	319	SerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGly 338
Db	850	GTCTATATTTTTCATCATCTAGAAATCC---TGTCAAAACTCTGATCGTTTGGAT----- 900
QY	339	LysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThr 358
Db	901	-----GAGGTATTTGAAATCCTCTTTAAGTCGTTCGAGAACTATATTCTGAACACA 951
QY	359	HisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLys----- 375
Db	952	TACAAAACAAAATTTTCACAGTTTCTGATGTTCTATGTCATGCTCCTACTAGATCCTGAAAAAT 1011
QY	376	-----LeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393
Db	1012	TGTGGTGTGAGATTTGCCAGTAAGCTGTTGGAC-----ATATATCTCTCC 1056
QY	394	SerAsnProAla---IleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAla 412
Db	1057	AGCAACACATGTCGACTTACTAGGATGAGTGAGTGGCTTATCTAGCTAGCTATTGTCT 1116
QY	413	ArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrp 432
Db	1117	CGTGGAAAGTTTTTGCCTGCTTCTTTTGTGGCTAGCATGTTGAAAAGATGGTGGACGAG 1176
QY	433	LeuHisIleTyrLeuAsn-----AsnGlnAspSerGlyThrLysAlaPheCysAspVal 450
Db	1177	TGTGCGGAATATTGCGGAAACATGCAATGATGATGTGAAGCCAGAAGCA----- 1224
QY	451	AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470
Db	1225	-----CATCAAGTGTCTATTCTGATGTCAGGCAATCTTGATGTGCTATGCTTCCGA 1278
QY	471	HisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSer----- 488
Db	1279	ATGAGATCCATAGTGGAG-----ATTCTCGCTTTCAATCGCAGTTT 1320
QY	489	LeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508
Db	1321	AGATCATTTGGAGTCAATTTTATCTCACAAACTAAACCCACTACTGGTGTGCTTCCATCT 1380
QY	509	ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528
Db	1381	GTAGTTTCCGAGTTCCTTAA-ACAAGCCAAAGCTGGTGGTCTGTTTCATTGCTCAGAATC 1439
QY	529	IleGlu-ArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSe 548
Db	1440	CTTCATTTTCGATGACCTACACGAGTCTGAGCTCTCTCGTGTCTTTGGTGGCTTTGAAAG 1499
QY	548	rValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLeuLy 568
Db	1500	G-----CTTGACACATCTTCCCGTTTGACCCGTCGCTTGTGAA 1538
QY	568	sArgSerLysLysPheIleAspProIleTyrGlnValTrp----- 581
Db	1539	AATGCTAGCAGCTACATCTCCCGAACTTCAACTTCTGGTCAATGGTGAAAAACGACTTA 1598

```
QY      582  ----GluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspI1 600
Db      1599  TGGAGAAGATGGTGACGAAGAGCTTTGTGATGAGGTCATAGTGAATGGAGATGCACACAG 1658
QY      600  eValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValI1 620
Db      1659  TGGCGAGGAACCTGACGATGACGTT-----GAACTTGATAGTGAGATGAACACGAT 1709
QY      620  eGlyIleThrPro 624
Db      1710  GTCTACAACTCCG 1722
```

Search completed: July 18, 2004, 14:02:24  
Job time : 818 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 08:30:12 ; Search time 126 Seconds  
(without alignments)  
2867.246 Million cell updates/sec

Title: US-09-932-678-2  
Perfect score: 3403  
Sequence: 1 MAAPLLHTRLPGDAAASSA.....RSPSSVGSPPVLYMQPSPL 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09932678/runat\_15072004\_103604\_5238/app\_query.fasta\_1.839  
-DB=Issued Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09932678 @CGN 1 1 69 @runat\_15072004\_103604\_5238 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCRTUS COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369	10.8	513	4	US-09-621-976-3086 Sequence 3086, Ap
2	138	4.1	263	4	US-09-313-294A-919 Sequence 919, App
3	132	3.9	262	4	US-09-313-294A-924 Sequence 924, App
4	111	3.3	4392	4	US-09-470-661A-3 Sequence 3, Appli
5	111	3.3	15413	4	US-09-981-282-2 Sequence 4, Appli
6	111	3.3	15424	4	US-09-981-282-2 Sequence 2, Appli
7	111	3.3	15424	4	US-10-143-186-2 Sequence 2, Appli
8	111	3.3	15450	4	US-09-470-661A-1 Sequence 1, Appli
9	106	3.1	15413	4	US-09-981-282-3 Sequence 3, Appli
10	106	3.1	15424	4	US-09-981-282-1 Sequence 1, Appli
11	106	3.1	15424	4	US-10-143-186-1 Sequence 1, Appli
12	104.5	3.1	580073	4	US-08-545-528D-1 Sequence 1, Appli

13	103.5	3.0	2115	1	US-08-329-681A-3	Sequence 3, Appli
14	103	3.0	8054	4	US-09-581-909-2	Sequence 2, Appli
15	102.5	3.0	1692	4	US-09-134-001C-335	Sequence 335, App
16	102	3.0	10968	3	US-09-228-246-1	Sequence 1, Appli
17	101	3.0	5475	2	US-08-680-327-1	Sequence 1, Appli
18	101	3.0	5475	3	US-09-228-246-3	Sequence 3, Appli
19	101	3.0	580073	4	US-08-545-528D-1	Sequence 1, Appli
20	100.5	3.0	4378	2	US-09-080-897-3	Sequence 3, Appli
21	100.5	3.0	4378	3	US-09-323-735-3	Sequence 3, Appli
22	100.5	3.0	4399	3	US-08-899-595-2	Sequence 2, Appli
23	100.5	3.0	6806	4	US-08-956-171E-472	Sequence 472, App
24	99.5	2.9	3024	4	US-09-957-005-8	Sequence 8, Appli
25	99.5	2.9	5361	3	US-08-973-462-2	Sequence 2, Appli
26	99.5	2.9	6152	3	US-08-973-462-1	Sequence 1, Appli
27	99	2.9	10323	4	US-09-280-428A-11	Sequence 11, Appli
28	98.5	2.9	2402	3	US-08-776-265-4	Sequence 4, Appli
29	98.5	2.9	2402	4	US-09-398-184-4	Sequence 4, Appli
30	98.5	2.9	4565	3	US-08-776-265-1	Sequence 1, Appli
31	98.5	2.9	4565	4	US-09-398-184-1	Sequence 1, Appli
32	98.5	2.9	5902	4	US-09-381-989-4	Sequence 4, Appli
33	98.5	2.9	89047	4	US-09-596-002-34	Sequence 34, Appli
34	97.5	2.9	4915	4	US-09-173-053-7	Sequence 7, Appli
35	97.5	2.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
36	97	2.9	1638	1	US-08-417-330A-13	Sequence 13, Appli
37	97	2.9	10968	2	US-08-680-327-2	Sequence 2, Appli
38	96.5	2.8	813	4	US-09-872-733A-19	Sequence 19, Appli
39	96.5	2.8	1360	4	US-09-082-649B-80	Sequence 80, Appli
40	96.5	2.8	1360	4	US-09-082-649B-81	Sequence 81, Appli
41	96.5	2.8	1738	4	US-09-574-873-9	Sequence 9, Appli
42	96.5	2.8	3418	2	US-08-944-916-12	Sequence 12, Appli
43	96.5	2.8	3547	4	US-09-340-798A-43	Sequence 43, Appli
44	96.5	2.8	3610	4	US-09-194-949A-1	Sequence 1, Appli
45	96.5	2.8	3987	4	US-09-082-649B-83	Sequence 83, Appli

ALIGNMENTS

RESULT 1

US-09-621-976-3086  
; Sequence 3086, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3086  
; LENGTH: 513  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 25..231  
US-09-621-976-3086

Alignment Scores:  
Pred. No.: 2.1e-36 Length: 513  
Score: 369.00 Matches: 75  
Percent Similarity: 88.64% Conservative: 3  
Best Local Similarity: 85.23% Mismatches: 5  
Query Match: 10.84% Indels: 6  
DB: 4 Gaps: 1

US-09-932-678-2 (1-651) x US-09-621-976-3086 (1-513)

Qy 528 ilelleGluArgAsnArgGlnMetLeuProValileArgSerThrAlaGlyGlyasp 547  
Db 1 ATCATTGAGAGGAACAATGCCAGATGCTGCCAGTATAGGAGTACCACTGGAGGAGAC 60



Best Local Similarity: 21.23%				Mismatches: 231			
Query Match: 3.26%				Indels: 178			
DB: 4				Gaps: 27			
US-09-932-678-2 (1-651) x US-09-470-661A-3 (1-4392)							
QY	2	AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerSerAlaVal	21				
Db	229	GCAGTTCCTTCGCTTATAGATGTCCTGATCTCCGGTGTGACGCATCTCCTAAGTTACTC	288				:::
QY	22	LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe	41	:::			
Db	289	GCTCGTCACGGCCCGGGAACACTGGGATCGATGGCACGCTTTGG-----GACTTT	339				
QY	42	PheAsnSerProProArgLysThrValArgPheGly-----	53				
Db	340	GAGGCCGAGGCCACCAAGAGGAAATTGCACCTCAGTCCGCAATAATACAGGCTTGTGAC	399				
QY	54	-----GlyThrValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAsp	70				
Db	400	ATTAGCGCGCGCAGCGCACCTGAAATTGGTCTCCCTTACAAG-----	441				
QY	71	PheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAsn	90				
Db	442	-----CTGTACCCTGTTAGGGCAACCCCTGAGCGGGTAAAA	477				
QY	91	TrpLeuLeuGlu-PheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuI1	110	:::			
Db	478	GGAGTTTACAGAAATACAAGTTTGGAGACATACCTTACAAAACCCCGAGTGACATGGA	537				
QY	110	eSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLe	130	:::			:::
Db	538	AGCCCACTGCACGGCGTGCCTGC-CTCAGGCCCAATGCCACTCCGGTGACTGAT	591				
QY	130	uAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMe	150				
Db	592	-----GGCGGCTGTCTTGGCTACTACCATG-----CCCTCCGGTTTGA	632				
QY	150	tIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAspValse	170	:::			
Db	633	ATTGTATGTACCGACCATTCAGCGTCTGTCTCTT-----GATTATCT	674				
QY	170	rAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe	190				
Db	675	TGACTCTAGGCCTGACTGCCCCAAACAGTTGACAGACGCGGCTGTGAGGATGCCGCATT	734				
QY	190	uGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetPro-----I1	207				
Db	735	GAGAGACCTCTCCAAGTATGACTTGTCCACCCCAAGGCTTGTTTTACCTGGGGTTCTTCG	794				
QY	207	eLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys---TyrVa	226				
Db	795	CCTTGTGCGTAAGTACCTGTTGCCCATGTGGGT-----AAGTCCCGCCCGCT	842				
QY	226	lHisAsnLeuLeuArgIleSerValTyr-----	235				
Db	843	TCAT-----CGGCCTTCCACTTACCCTGCCAAGAATTCTATGGCTGGAATAAATGG	893				
QY	236	-----PheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLy	253				
Db	894	GAACAGGTTTCCAACCAAGGAC-----ATTACAGCGCTCCCTGA	932				
QY	253	sLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr----	271				
Db	933	AATCGACGTTCTGTGCGCA---CAGGCGGTGCGAGAAAACTGGCAAACTGTACCCCTTG	989				
QY	272	-----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAs	287				
Db	990	TACCCTCAAGAAACAGTATTGTGG-----	1014				
QY	287	pGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetVa	307				
Db	1015	-----AAGAAGAGACTAGGACAAATACTCGGCACCAATAATTTCATTGCGTTGGC	1064				

QY	307	lHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAs	327				:::
Db	1065	CCACCGGCAGCG-----TTGAGTGGTGTACCCAGGGCTTCATGAAAAA	1109				
QY	327	pValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLe	347	:::			
Db	1110	GGCG-----TTTAACTCGCCCATCGCCCTCGGGAACAA	1146				
QY	347	uIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePh	367				
Db	1147	-----TTTAAGGAGCTACAGACTCCGGTCTTAGGCAGGTGC-----	1182				
QY	367	eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTr	387				
Db	1183	-----CTTGAAGCTGATCTTGCATCCTGT--	1206				
QY	387	pLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrI1	407				
Db	1207	-----GATCGATCCACACACTGCAATTGTCGGCTGGTTTGGCGCAATCTTCT	1253				
QY	407	eGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAs	427	:::			
Db	1254	TTATGAACCTTGCTGTGCTGAAGAGCACCTACCGTCTGCTGCTGAACCTGCTGCCATGA	1313				
QY	427	pLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPh	447				
Db	1314	CCTATTGGTC-----ACGAGTCCGGCGCAGTGACTAAGAG	1349				
QY	447	eCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh	467	:::			
Db	1350	GGTGGCTATCGTCTGGCGACCCGATCACTTCTGTGCTAAACACCACTTACAGCTTGT	1409				
QY	467	eValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGlyLeuG1	484	:::			
Db	1410	GATATATGCACAGCACATGGTGTCTTAGTTACTTTAAAAAGTGGTCACCCCTCATGGCCTTCT	1469				
QY	484	nTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLy	503	:::			
Db	1470	GTTCCTACAAGACCAGCTGAAGTTCGAGGAC-----ATGCTCAAAGTCCAACCCCTGAT	1523				
QY	503	sIleCys-----LeuProSerValValAsnPhePhe--	513				
Db	1524	CGTCTATTCCGACGACCTCGTGTGTATGCCGAATCTCCACCATGCCGAACCTACCACTG	1583				
QY	514	-----AlaAlaI1	516				
Db	1584	GTGGGTGGAACATCTGAATTGTATGTGGTGTTCAGACGGACCCAAAGACAGCCCAT	1643				
QY	516	eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMe	536				
Db	1644	AACGGACTGCCATCATTTCTAGGCTGTAGGATAATA-----AATGGACGCCAGCT	1694				
QY	536	tLeuPro	538	:::			
Db	1695	AGTCCCC	1701				

RESULT 5

US-09-981-282-4  
; Sequence 4, Application US/09981282  
; Patent No. 6641819  
; GENERAL INFORMATION:  
; APPLICANT: Mengeling, William L  
; APPLICANT: Vorwald, Ann  
; APPLICANT: Lager, Kelly  
; APPLICANT: Roof, Mike  
; APPLICANT: Burkhardt, Kelly  
; APPLICANT: Gorcycza, David  
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE  
; FILE OF INVENTION: BASED ON ISOLATE JA-142  
; FILE REFERENCE: 27093B  
; CURRENT APPLICATION NUMBER: US/09/981,282  
; CURRENT FILING DATE: 2002-08-19

```
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 15413
; TYPE: RNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
; US-09-981-282-4

Alignment Scores:
Pred. No.:      0.124      Length:      15413
Score:          111.00     Matches:      86
Percent Similarity: 32.92%  Conservative: 47
Best Local Similarity: 21.29% Mismatches:    137
Query Match:      3.26%    Indels:       134
DB:              4        Gaps:         21

US-09-932-678-2 (1-651) x US-09-981-282-4 (1-15413)

QY 189 AlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetPro----- 206
Db 8411 GCAUUAAGAGAGACCUCCUCCAGUAUGACUUGUCCACCCCAAGGCUUUGUUUACCUUGGAGUU 8470

QY 207 ---IleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys--- 224
Db 8471 CUUCGCCUUGUGCGUAAGUACCUUGUUGUCUAGUGGGU-----AAGUGCCCG 8518

QY 225 TyrValHisAsnLeuLeuArgIleSerValTyr----- 235
Db 8519 CCCGUUCAU-----CGGCCUUCACUACUACCCUGCCCAAGAAUUCUUAUGGCUUGGAAUA 8569

QY 236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeu 251
Db 8570 AAUGGGAACAGGUUCCACCAAGGAC-----AUCCAGAGCGUC 8608

QY 252 LeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr 271
Db 8609 CCUGAAAUAGCAGCUUCUGUGCGCA---CAGGCCUGGCGGAAACUUGGCAAAUCUGUUAAC 8665

QY 272 -----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285
Db 8666 CCUUGUACCCUCAGAAACAGUAUUGGG----- 8695

QY 286 MetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGln 305
Db 8696 -----AAGAAGAGACUAGGACAAUACUCGGCACCAAAUACUUAUUGCG 8740

QY 306 MetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMet 325
Db 8741 CUGGCCACCGGGCAGCG-----UUGAGUGGUGUACCCAGGCGUUCUUG 8785

QY 326 LysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArg 345
Db 8786 AAAAAGGCG-----UUUAAUCUGCCCAUUGCCUUGGUAACAA----- 8827

QY 346 AspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGln 365
Db 8828 -----UUUAAAGACCUUCAGACUCCGUCUUAAGCAGGUGC----- 8863

QY 366 PhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis 385
Db 8864 -----CUUGAAGCUGAUUCUUGCAUCC 8884

QY 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn 405
Db 8885 UGC-----GAUCGUCCACACACCUGCAAUUGUCCGUGGUUGCGCCCAAU 8929

QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
Db 8930 CUUCUUUAUGAACUUGCCUGUGCAGAGACACCGCCGUCGUGUAGAACUCGUCG 8989

QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445
Db 8990 CAGGACCUACUGGUC-----ACGAGUCCGGCGAGUAACU 9025

; 446 AlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyr 465
; 9026 AAGAGAGGUGGCCUGUCUGGCGGACCCGAUCACUUCUGUGUCCACCAUUUACAGC 9085
; 466 ThrPheValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGly 482
; 9086 UUGGUGAUUAUUGCACAACACAUUGGUGCUCAGUUAUUUAAAAGUGGUCACCCUCAUGGC 9145
; 483 LeuGlnTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501
; 9146 CUUCGUUUUCUACAGACACGACGUGAAGUUUGAGGAC-----AUGCUCAAGGUUACACCC 9199
; 502 LeuLysIleCys-----LeuProSerValValAsnPhe 512
; 9200 CUGAUGGUCUAUUCGAGACGACCUUGUACUGUAUGGCGGAGUCUCCCAACCAUAC 9259
; 513 Phe-----Ala 514
; 9260 CACUGGUGGUUGAACAUCUGAACCUUGAUGGUGGUUUUCAGACGGACCCAAAGAGACA 9319
; 515 AlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArg 534
; 9320 GCCAUAAACAGACUCGCCCAUCAUUCUAGGCUUGAGGAUAUA-----AAUGGACGC 9370

; 535 GlnMetLeuPro 538
; 9371 CAGCUGGUCUCCU 9382

RESULT 6
US-09-981-282-2
; Sequence 2, Application US/09981282
; Patent No. 6641819
; GENERAL INFORMATION:
; APPLICANT: Mengeling, William L
; APPLICANT: Vorwald, Ann
; APPLICANT: Lager, Kelly
; APPLICANT: Roof, Mike
; APPLICANT: Burkhardt, Kelly
; APPLICANT: Gorcyca, David
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE
; TITLE OF INVENTION: BASED ON ISOLATE JA-142
; FILE REFERENCE: 27093B
; CURRENT APPLICATION NUMBER: US/09/981,282
; CURRENT FILING DATE: 2002-08-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 15424
; TYPE: DNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-09-981-282-2

Alignment Scores:
Pred. No.:      0.125      Length:      15424
Score:          111.00     Matches:      86
Percent Similarity: 32.92%  Conservative: 47
Best Local Similarity: 21.29% Mismatches:    137
Query Match:      3.26%    Indels:       134
DB:              4        Gaps:         21

US-09-932-678-2 (1-651) x US-09-981-282-2 (1-15424)

QY 189 AlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetPro----- 206
Db 8412 GCATTAAAGAGACCTCTCCAAAGTATGACTTGTCACCCACCAAGGCTTGTTTTACCTGGAGTT 8471

QY 207 ---IleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys--- 224
Db 8472 CTTCGCCCTTGCGTAAGTACTGTTTGCTCATGTGGGT-----AAGTGCCCG 8519

QY 225 TyrValHisAsnLeuLeuArgIleSerValTyr----- 235
Db 9025 CAGGACCUACUGGUC-----ACGAGUCCGGCGAGUAACU 9025
```

Db 8520 CCCGTTTCAT-----CGGCCTTCCACTTACCCTGCCAAGAAATTCATGCTGGAATA 8570

QY 236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeu 251

Db 8571 AATGGGAACAGGTTTCCAACCAAGGAC-----ATCCAGAGCGTC 8609

QY 252 LeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr 271

Db 8610 CCTGAAATCGAGCTTCTGTGCGCA--CAGGCCGTGCGGGAATACTGGCAAACTGTTACC 8666

QY 272 -----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285

Db 8667 CCTGTACCCTCAAGAAACAGTATTGTGG-----8696

QY 286 MetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGln 305

Db 8697 -----AAGAAGAAGACTAGGACAATACTCGSCACCAATAACTTCATTGCG 8741

QY 306 MetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMet 325

Db 8742 CTGGCCCAACCGGCAGCG-----TTGAGTGGTGTCAACCCAGGGCTTCATG 8786

QY 326 LysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArg 345

Db 8787 AAAAAGCG-----TTTAACTCGCCCATTGCCCTCGGTAAACAA-----8828

QY 346 AspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGln 365

Db 8829 -----TTTAAAGAGCTTCAGACTCGGTCCTTAGGCAGGTGC-----8864

QY 366 PhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis 385

Db 8865 -----CTGAAGCTGATCTTGCAATCC 8885

QY 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn 405

Db 8886 TGC-----GATCGCTCCACACCTGCAATTGTCGCTGTTGCCGCCAAT 8930

QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425

Db 8931 CTTCTTTATGAACTTGCCCTGTGCTGAAGAGCACCTGCCGTGCTACGTGTGAACGTGCTGC 8990

QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445

Db 8991 CACGACCTACTGGTC-----ACGCAGTCCGCGCGCAGTAAC 9026

QY 446 AlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyr 465

Db 9027 AAGAGAGGTGGCCTGTCTGCTGGCGACCCGATCACTTCTGTGTCCAACACCATTTACAGC 9086

QY 466 ThrPheValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGly 482

Db 9087 TTGGTGATATATGCACAACATGGTGTCTCAGTTACTTTAAAGTGGTCAACCTCATGGC 9146

QY 483 LeuGlnTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501

Db 9147 CTTCTGTTTCTACAAGACCAGCTGAAGTTTGAGGAC-----ATGCTCAAGGTTCAACCC 9200

QY 502 LeuLysIleCys-----LeuProSerValValAsnPhe 512

Db 9201 CTGATGCTATATTCGGACGACCTCGTACTGTATGCCGAGTCTCCACCATGCCAAACTAC 9260

QY 513 Phe-----Ala 514

Db 9261 CACTGGTGGTTGAACATCTGAACCTGATGCTGGGTTTTTCAGACGGACCCAAAGAAGACA 9320

QY 515 AlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArg 534

Db 9321 GCCATAACAGACTCGCCATCATTTCTAGGCTGTAGGATAATA-----AATGGACGC 9371

QY 535 GlnMetLeuPro 538

Db 9372 CAGCTCGTCCCT 9383

RESULT 7

US-10-143-186-2

; Sequence 2, Application US/10143186

; Patent No. 6660513

; GENERAL INFORMATION:

; APPLICANT: MENGELING, WILLIAMS L.

; APPLICANT: VORWALD, ANN

; APPLICANT: LAGER, KELLY

; APPLICANT: ROOF, MIKE

; APPLICANT: BURKHART, KELLY

; APPLICANT: GORCYCA, DAVID E

; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE,

; TITLE OF INVENTION: BASED ON ISOLATE JA-142

; FILE REFERENCE: 27093a

; CURRENT APPLICATION NUMBER: US/10/143,186

; CURRENT FILING DATE: 2002-05-09

; PRIOR APPLICATION NUMBER: US/09/461,879

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: 09/298,110

; PRIOR FILING DATE: 1999-04-22

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 15424

; TYPE: DNA

; ORGANISM: Porcine reproductive and respiratory syndrome virus

US-10-143-186-2

Alignment Scores:

Pred. No.: 0.125 Length: 15424

Score: 111.00 Matches: 86

Percent Similarity: 32.92% Conservative: 47

Best Local Similarity: 21.29% Mismatches: 137

Query Match: 3.26% Indels: 134

DB: 4 Gaps: 21

US-09-932-678-2 (1-651) x US-10-143-186-2 (1-15424)

QY 189 AlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetPro----- 206

Db 8412 GCATTAAAGAGACCTCTCCAAGTATGACTGTCTCCACCCCAAGGCTTTGTTTACCTGGAGTT 8471

QY 207 ---IleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys--- 224

Db 8472 CTTGCGCTTGTGCGTAAGTACCTGTTTGCTCATGTGGGT-----AAGTGCCCG 8519

QY 225 TyrValHisAsnLeuLeuArgIleSerValTyr----- 235

Db 8520 CCCGTTTCAT-----CGGCCTTCCACTTACCTGCCAAGAATTCATGGCTGGAATA 8570

QY 236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeu 251

Db 8571 AATGGGAACAGGTTTCCAACCAAGGAC-----ATCCAGAGCGTC 8609

QY 252 LeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr 271

Db 8610 CCTGAAATCGACGTTCTGTGCGCA--CAGGCCGTGCGGGAATACTGGCAAACTGTTACC 8666

QY 272 -----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285

Db 8667 CTTGTACCTCAAGAAACAGTATTGTGG-----8696

QY 286 MetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGln 305

Db 8697 -----AAGAAGAAGACTAGGACAATACTCGGCACCAATAACTTCATTGCG 8741

QY 306 MetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMet 325

Db 8742 CTGGCCCAACCGGCAGCG-----TTGAGTGGTGTCAACCCAGGGCTTCATG 8786

QY 326 LysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArg 345





```
Db      8653  TACCCTCAAGAAACACGATTGTGGG----- 8677
QY      287  pGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetVa 307
Db      8678  -----AAGAAGAAGACTAGGACAATACTCGGCACCAATAATTTCATTGCGTTGGC 8727
QY      307  lHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAs 327
Db      8728  CCACCGGCGCAGCG-----TTGAGTGGTGTCAACCCAGGCTTCATGAATAA 8772
QY      327  pValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLe 347
Db      8773  GGCG-----TTTAACTCGCCCATCGCCCTCGGGAATAA----- 8809
QY      347  uIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
Db      8810  -----TTTAAGGAGCTACAGACTCCGGTCTTAGGCAGGTGC----- 8845
QY      367  eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTr 387
Db      8846  -----CTTGAAGCTGATCTTGTCATCCTGT-- 8869
QY      387  pLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIl 407
Db      8870  -----GATCGATCCACACCTGCAATTGTCCGCTGTTTGGCCCAATCTTCT 8916
QY      407  eGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAs 427
Db      8917  TTATGAACCTGCCTGTGCTGAAGAGCAGCCTACCGTGGTACGTGCTGAACCTGCTGCCATGA 8976
QY      427  pLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPh 447
Db      8977  CCTATTGGTC-----ACGCACTCCGGCGCAGTGACTAAGAG 9012
QY      447  eCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
Db      9013  GGGTGGCTATCGTCTGGCGACCCGATCACTTCTGTGTCTAACCACTTACAGCTTGGT 9072
QY      467  eValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGlyLeuGl 484
Db      9073  GATATATGCACAGCACATGGTGTCTTAGTTACTTTAAAGTGGTCACCCCTCATGGCCTTCT 9132
QY      484  nTyrLeuGln--SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLy 503
Db      9133  GTTCCTACAAGACCAGCTGAAGTTCGAGGAC-----ATGCTCAAAGTCCAACCCCTGAT 9186
QY      503  sIleCys-----LeuProSerValValAsnPhePhe-- 513
Db      9187  CGTCTATTCCGACGACCTCGTGTGTATGCGGAATCTCCACCATGCGCGAATCACTACCATG 9246
QY      514  -----AlaAlaIl 516
Db      9247  GTGGGTGGAACATCTGAATTTGATGCTGGGTTTTCAGACGAGACCCAAAGAAGACAGCCAT 9306
QY      516  eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMe 536
Db      9307  AACGGACTCGCCATCATTTCTAGGCTGTAGGATAATA-----AATGGACGCCAGCT 9357
QY      536  tLeuPro 538
Db      9358  AGTCCCC 9364
```

RESULT 9

```
US-09-981-282-3
; Sequence 3, Application US/09981282
; Patent No. 6641819
; GENERAL INFORMATION:
; APPLICANT: Mengeling, William L
; APPLICANT: Vorwald, Ann
; APPLICANT: Lager, Kelly
; APPLICANT: Roof, Mike
; APPLICANT: Burkhardt, Kelly
```

```
; APPLICANT: Gorcyca, David
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE
; TITLE OF INVENTION: BASED ON ISOLATE JA-142
; FILE REFERENCE: 27093B
; CURRENT APPLICATION NUMBER: US/09/981,282
; CURRENT FILING DATE: 2002-08-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 15413
; TYPE: RNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-09-981-282-3

Alignment Scores:
Pred. No.: 0.524 Length: 15413
Score: 106.00 Matches: 127
Percent Similarity: 32.17% Conservative: 67
Best Local Similarity: 21.06% Mismatches: 232
Query Match: 3.11% Indels: 178
DB: 4 Gaps: 27
```

```
US-09-932-678-2 (1-651) x US-09-981-282-3 (1-15413)

QY      2  AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAlaVal 21
Db      7910  GCAGUCCUUCGCUUAAGACGCUUAUACUCCGGCGUGAUGCAUCCCAAGUUACUC 7969
QY      22  LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe 41
Db      7970  GCGCGCACGGGCGGGAACACACUGGAUCCGUGGACCGCUUUGG-----GAUUUU 8020
QY      42  PheAsnSerProProArgLysThrValArgPheGly----- 53
Db      8021  GAGCGCGAGGCCACUAAAGAGGAAAUUGCACUCAGUGCGCAAAUAUACAGGCUUGAC 8080
QY      54  -----GlyThrValThrGluValLeuLeuLysTyrLysGlyGluThrAsnAsp 70
Db      8081  AUUAGCGCGCGCGACGCGACCUAGAAUUGGUCUCCUUAUAG----- 8122
QY      71  PheGluLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAsn 90
Db      8123  -----CUGUACCCUGUCAGGGCAACCCUGAGCGGGUAAAA 8158
QY      91  TrpLeuLeuGlu-PheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIl 110
Db      8159  GGAGUUUUACAGAAUACAAGUUUUGGAGAUUACCUUAUAAACCCCGACAGACACUGGA 8218
QY      110  eSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLe 130
Db      8219  AGCCACAGUCACGCGGCGUGCCUGC-CUCACGCCCAUUGCCACUCCGGUGACUGAU----- 8272
QY      130  uAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMe 150
Db      8273  -----GGGCGCUCGUCUUGGCCACGACUAUG-----CCUCCGGUUUUGA 8313
QY      150  tIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValaspValse 170
Db      8314  GUUGUUGUACCGACCAUUCACGCGUCUGUCCUU-----GAUUAUCU 8355
QY      170  rAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190
Db      8356  UGAUUUCUAGGCCUGACUGCCCCAAACAGUUUGACAGACGCGCUGUGAGGACGCCGAUU 8415
QY      190  uGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetPro-----Il 207
Db      8416  AAGAGACCUCCUCCAAGUAUGACUUGUCCACCCCAAGGCUUUUUUUAACCGAGGUUCUCG 8475
QY      207  eLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys---TyrVa 226
Db      8476  CCUUGUGCGUAAAGUACCUUGUUUGCUCAUGUGGGU-----AAGUGCCCGCCCGU 8523
QY      226  lHisAsnLeuLeuArgIleSerValTyr----- 235
```

Db 8524 UCAU-----CGGCCUCCACUUAACCCGCAAGAAUUCUAUGGCGGAUAAUAAUGG 8574  
QY 236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLy 253  
Db 8575 GAACAGGUUCCAAACAGGAC-----AUCCAGAGCGUCCCUCA 8613  
QY 253 sLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr----- 271  
Db 8614 AAUCGAGUUCUGUGCGCA--CAGGCCGUUCGGGAAACUGGCAACUUCUACCCCUUG 8670  
QY 272 -----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAs 287  
Db 8671 UACCCUCAAGAAACAGUAUUGUGG----- 8695  
QY 287 pGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetVa 307  
Db 8696 -----AAGAAGAGACUAGGACAAUACUCGGCACCAAAUACUUAUUGCGUGGC 8745  
QY 307 lHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAs 327  
Db 8746 UCACCGGCGACG-----UUGAGUGUGUCACCCAGGCGUUCAUUGAAGAAA 8790  
QY 327 pValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLe 347  
Db 8791 GGCG-----UUUAAUCGCGCCAUUGCCUCGUAACAAA----- 8827  
QY 347 uIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367  
Db 8828 -----UUUAAAGAGCUUCAGACUCCGGUCUUAAGGCGAGGUC----- 8863  
QY 367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTr 387  
Db 8864 -----CUUGAAGCUGAUCUUGCAUCCUGC-- 8887  
QY 387 pLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIl 407  
Db 8888 -----GAUCGCUCCACACCGCAUUGCGCGUGGUUUGCGCCAAUUCU 8934  
QY 407 eGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAs 427  
Db 8935 UUAUGAACUUGCCUGUGCUGAAGAGCACCAGCCGUGUACGUGUUGAUAACUGCGCACGA 8994  
QY 427 pLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPh 447  
Db 8995 CCUACUGGUC-----ACGCAGUCCGCGCAGUAACUUAAGAG 9030  
QY 447 eCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467  
Db 9031 AGGUGGCCUGUGUGUGCGACCCGAUCACUUCUGUGUCCAACACCAUUUACAGCUUGU 9090  
QY 467 eValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGlyLeuGl 484  
Db 9091 GAUAUAUGCACAAACAUGGUGGUCAGUAUUUAAAAAGUGGUCACCCUUAUGGCCUUCU 9150  
QY 484 nTyrLeuGln----SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLy 503  
Db 9151 GUUUCUACAAGACCAGCUGAAGUUGAGGAC-----AUGCUAAGGUUCAACCCUUGAU 9204  
QY 503 sIleCys-----LeuProSerValValAsnPhePhe-- 513  
Db 9205 CGUCUAUUCGGACGACCUUGUACUGUAUGCCGAGUCUCCACCAUGCAACUACCAUG 9264  
QY 514 -----AlaAlaIl 516  
Db 9265 GUGGGUUGAACAUUCUGAACCUUGAUGCGGUUUUUCAGACGGACCCAAAGAACAGCCAU 9324  
QY 516 eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnArgGlnMe 536  
Db 9325 AACAGACUGCCCAUCAUUUCUAGGCGUGAGGAUAUA-----AAUGGACGCCAGCU 9375  
QY 536 tLeuPro 538  
: : : : : | | |

Db 9376 CGUCCCU 9382  
RESULT 10  
US-09-981-282-1  
; Sequence 1, Application US/09981282  
; Patent No. 6641819  
; GENERAL INFORMATION:  
; APPLICANT: Mengeling, William L  
; APPLICANT: Vorwald, Ann  
; APPLICANT: Lager, Kelly  
; APPLICANT: Roof, Mike  
; APPLICANT: Burkhardt, Kelly  
; APPLICANT: Gorcyca, David  
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE  
; FILE REFERENCE: 27093B  
; CURRENT APPLICATION NUMBER: US/09/981,282  
; CURRENT FILING DATE: 2002-08-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 15424  
; TYPE: DNA  
; ORGANISM: Porcine reproductive and respiratory syndrome virus  
US-09-981-282-1  
  
Alignment Scores:  
Pred. No.: 0.525 Length: 15424  
Score: 106.00 Matches: 127  
Percent Similarity: 32.17% Conservative: 67  
Best Local Similarity: 21.06% Mismatches: 232  
Query Match: 3.11% Indels: 178  
DB: 4 Gaps: 27  
  
US-09-932-678-2 (1-651) x US-09-981-282-1 (1-15424)  
QY 2 AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAlaVal 21  
Db 7911 GCAGTTCCTTCGCTTATAGACGTCTTAATCTCCGGCGTGTATGCATCTCCCAAGTTACTC 7970  
QY 22 LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe 41  
Db 7971 GCCGCCACGGCGCGGAAACACTGGGATCGATGGCACGCTTTGG-----GATTTT 8021  
QY 42 PheAsnSerProProArgLysThrValArgPheGly----- 53  
Db 8022 GAGGCCGAGGCCACTAAAGAGGAAATTCACCTCAGTCGCGCAATAATACAGGCTTTGTGAC 8081  
QY 54 -----GlyThrValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAsp 70  
Db 8082 ATTAGCGCGCGGACGACCTGAAATTTGGTCTTCTTATAAG----- 8123  
QY 71 PheGluLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAsn 90  
Db 8124 -----CTGTACCTGTCTCAGGGGCAACCTGAGCGGGTAAAA 8159  
QY 91 TrpLeuLeuGlu-PheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIl 110  
Db 8160 GGAGTTTACAGAATAACAAGGTTGGAGATATACCTTATAAAACCCCGAGTGCACCTGGA 8219  
QY 110 eSerIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLe 130  
Db 8220 AGCCCAAGTGCAGCGGCTGCCTGC-CTCAGCGCAATGCCACTCCGGTGACTGAT---- 8273  
QY 130 uAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMe 150  
Db 8274 -----GGCGCTCCGCTTGGCCACGACTATG-----CCCTCCGGTTTGA 8314  
QY 150 tIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAspValSe 170  
Db 8315 GTTGATGTACCGACCATTCACGCGTCTGTCTT-----GATTATCT 8356  
QY 170 rAspSerAspAspGluAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190



Db 8160 GGAGTTTACAGAAATACAAGGTTTGGAGATATACCTTATAAAACCCCCAGTGACACTGGA 8219  
QY 110 eSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLe 130  
Db 8220 AGCCAGTGACGCGGTGCCTGC-CTCAGCGCCCAATGCCACTCCGGTGACTGAT----- 8273  
QY 130 uAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMe 150  
Db 8274 -----GGCGGTCCGTCTTGGCCACGACTATG-----CCCTCCGGTTTGA 8314  
QY 150 tIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValaspValse 170  
Db 8315 GTTGATGTACCGACCATCCAGCGTCTGCCTT-----GATTATCT 8356  
QY 170 rAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190  
Db 8357 TGATTCTAGGCCTGACTGCCCCAAACAGATTGACAGACACGGCTGTGAGGACGCCGCATT 8416  
QY 190 uGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetPro-----Il 207  
Db 8417 AAGAGACCTCTCCAAGTATGACTTGTCCACCCCAAGGCTTGTTTACCTGGAGTTCCTCG 8476  
QY 207 eLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys---TyrVa 226  
Db 8477 CCTTGTGCGTAAGTACCTGTTTGTCTCATGTGGT-----AAGTGCCCGCCCGT 8524  
QY 226 lHisAsnLeuLeuArgIleSerValTyr----- 235  
Db 8525 TCAT-----CGGCTTCCACTTACCTGCCAAGAATTCTATGGTGGAATAAATGG 8575  
QY 236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLy 253  
Db 8576 GAACAGGTTTCCAAACCAAGGAC-----ATCCAGAGCGTCCCTGA 8614  
QY 253 sLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr----- 271  
Db 8615 AATCGAGCTTCTGTGCGCA---CAGGCGTTCGGGAAAACTGGCAAACTGTTACCCCTTG 8671  
QY 272 -----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAs 287  
Db 8672 TACCCTCAAGAAACAGTATTGTGG----- 8696  
QY 287 pGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetVa 307  
Db 8697 -----AAGAAGAAGACTAGGACAACTACTCGGCACCAATACTTCATTGCGCTGGC 8746  
QY 307 lHisProValAlaGluArgLeuAspIleLeuMetSerSerLeuValLeuSerTyrMetLysAs 327  
Db 8747 TCACCGGCGCAGC-----TTGAGTGGTGTCACCCAGGCTTCATGAAAAA 8791  
QY 327 pValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgaspLe 347  
Db 8792 GGCG-----TTTAACTCGCCCATTTGCCCTCGGTAAAAACAA----- 8828  
QY 347 uIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367  
Db 8829 -----TTTAAAGAGCTTCAGACTCCGGTCTTAGCAGGTTGC----- 8864  
QY 367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTr 387  
Db 8865 -----CTTGAAGCTGATCTTGATCCTGC--- 8888  
QY 387 pLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIl 407  
Db 8889 -----GATCGTCCACACCTGCAATTGTCCGCTGGTTTGCGCCAATCTTCT 8935  
QY 407 eGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAs 427  
Db 8936 TTATGAAGCTTGCCTGTGCTGAAGAGCAGCAGCCGCTCGTACGTTGTGAAGTCTGCTGCCAGA 8995  
QY 427 pLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPh 447

Db 8996 CCTACTGGTCT-----ACGCAGTCCGGCGCAGTAACCTAAGAG 9031  
QY 447 eCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467  
Db 9032 AGGTGGCTGTCTGTGGCGACCCGATCACTTCTGTGTCCAACACCATTTACAGCTTGGT 9091  
QY 467 eValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGlyLeuGl 484  
Db 9092 GATATATGCACAACACATGGTGTCTAGTTACTTTTAAAAAGTGGTCAACCTCATGGCCTTCT 9151  
QY 484 nTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLy 503  
Db 9152 GTTTCTACAAGACCAAGCTGAAGTTTCAGGAC-----ATGCTCAAGGTTCAACCCCTGAT 9205  
QY 503 sIleCys-----LeuProSerValValAsnPhePhe-- 513  
Db 9206 CGTCTATTTCGGACGACCTCGTACTGTATGCCGAGTCTCCACCATGCCAAACTACCACTG 9265  
QY 514 -----AlaAlaIl 516  
Db 9266 GTGGGTTGAACATCTGAACCTGATGCTGGGTTTTCAGACGGACCCCAAGAGACAGCCAT 9325  
QY 516 eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMe 536  
Db 9326 AACAGACTCGCCATCATTTCTAGGCTGTAGGATAATA-----AATGGACGCCAGCT 9376  
QY 536 tLeuPro 538  
Db 9377 CGTCCCT 9383  
RESULT 12  
US-08-545-528D-1/c  
; Sequence 1, Application US/08545528D  
; Patent No. 6537773  
; GENERAL INFORMATION:  
; APPLICANT: Fraser et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment  
; Patent No. 6537773  
; FILE OF INVENTION: Thereof, and Uses Thereof  
; FILE REFERENCE: PB193P1  
; CURRENT APPLICATION NUMBER: US/08/545,528D  
; CURRENT FILING DATE: 1995-10-19  
; PRIOR APPLICATION NUMBER: US 08/488,018  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/473,545  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 580073  
; TYPE: DNA  
; ORGANISM: Mycoplasma genitalium  
US-08-545-528D-1  
Alignment Scores:  
Pred. No.: 409 Length: 580073  
Score: 104.50 Matches: 80  
Percent Similarity: 35.59% Conservative: 67  
Best Local Similarity: 19.37% Mismatches: 149  
Query Match: 3.07% Indels: 117  
DB: 4 Gaps: 18  
US-09-932-678-2 (1-651) x US-08-545-528D-1 (1-580073)  
QY 95 pheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleLeu 114  
Db 501760 TTTGATATTAAAGTTAAGTTATTAAACCCATCATTAAAGGTTTTTAATTCAAAAGGTAAAA 501701  
QY 115 ArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGly 134  
Db 501700 CAAATGGAA---TTAATGAAGAATAAAGTAATTAAACAGTCAAGATCTCTTAAACATTCAT 501644  
QY 135 AsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHis 154

Db 501643 GAATTGATTAATAACATTACCATTAATAATTAAAGGATTGTACCAATTTTCTT----- 501593

Qy 155 PheValProProArgValIleIleLysGluGlyAspValAspValSerAspSerAsp 174

Db 501593 ----- 501593

Qy 175 GluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAla 194

Db 501592 -----AATAACTTA-----AATTAAATGCTATGCGAAGGAATATGGATGAAGTTCAA 501545

Qy 195 ArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPhe 214

Db 501544 GAGGCGATTAGTTGACTT-----TTTCAAGCA 501518

Qy 215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeu-----LeuArgIle 232

Db 501517 GTTAGTTCCAAAGAAAAGCATTAAATTTAGTGGTTGAAAATATTGATGATTTTCGTAAT 501458

Qy 233 SerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeu 252

Db 501457 CAAATTCAACAATATGAAAAGAAAATAGACTTTTAAAAACAACAATT---AAAGCCATT 501401

Qy 253 LysLeuAspValAsnAlaSerArgGlnGlnGlyIleGluAspAlaGluGluThrAlaThrGln 272

Db 501400 GAATTA---GTGTTCCAAAATGAAACAGATGTGCATGATTAATTATTGAGCTTAATGAG 501344

Qy 273 ThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThr 292

Db 501343 AACTGCGATCTAATT-----ACTAAAAATATTAATACTTAAACCATCAAAAGTTTAGCT 501290

Qy 293 GluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGlu 312

Db 501289 CAGAATCAAATCAAT-----TATGAA 501269

Qy 313 ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValasp 332

Db 501268 AAACGTGTTTACTTGTAGTCAACAAACAACTCAAGCTTTAGAA----- 501227

Qy 333 GlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAsp 352

Db 501226 ---AAACTGAAAAATTCATAACT---GATTACTTAATTTAGCAGTTAATAAGTTTGAT 501173

Qy 353 LysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCys 372

Db 501172 -----GACTATCGCTTTTGTCTATACTTTTAGAT 501143

Qy 373 SerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGlnAsp 392

Db 501142 GATTTCCGATTTAAATTTTTCAG-----ATTGAAAAATTAAATTAGCGAT---CAAGAT 501092

Qy 393 ProSerAsnProAla----- 397

Db 501091 TTAAAGGTTCTCTGCTAAACATTTGCAAGTAATTAATAATCGAAAAATTAGAACTTTGATAAC 501032

Qy 398 IleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIle 417

Db 501031 ATCTTCACCTGAAGCTAATAATAAATACTATTAGGTACTTTTAATTTGTTTAGTGAAAAGATA 500972

Qy 418 ProLeuIleThrValLysSerCysLeuAspLeuValAsnTyrPheLeuHisIleTyrLeu 437

Db 500971 GATTTATTGCAAAATTCAAATTG----- 500951

Qy 438 AsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyr 457

Db 500950 -----ACTGAAGTTATTGTGTGATGTTGTTAGTTTAGTTAACTTAAGG 500909

Qy 458 SerAlaCys---GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSer 476

Db 500908 CTAATGTGCAAAACAAGCTTTTCTTTTTCGCAATAAATAATCGTCAAAAG-----AGT 500858

Qy 477 GlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeu 489

Db 500857 AGTCATATCGATAGATCATTAGAACACAGTTAACAAAACTA 500819

RESULT 13

US-08-329-681A-3

; Sequence 3, Application US/08329681A

; Patent No. 5650294

; GENERAL INFORMATION:

; APPLICANT: Kurth, Roland

; APPLICANT: Philippsen, Peter

; APPLICANT: Steiner, Sabine

; APPLICANT: Wright, Martin C.

; TITLE OF INVENTION: No. 5650294el Promoter Region

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Keil & Weinkauff

; STREET: 1101 Connecticut Avenue

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette-3.5 inch, 1.4 Mb storage

; COMPUTER: IBM AT-compatible, 80486 processor

; OPERATING SYSTEM: MS-DOS version 6.0

; SOFTWARE: WordPerfect version 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/329,681A

; FILING DATE: 26-OCT-1994

; CLASSIFICATION: 435

; CLASSIFICATION: C12P 25/00

; CLASSIFICATION: C12N 1/14

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2115 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

US-08-329-681A-3

Alignment Scores:

Pred. No.: 0.0356 Length: 2115

Score: 103.50 Matches: 50

Percent Similarity: 40.54% Conservative: 25

Best Local Similarity: 27.03% Mismatches: 71

Query Match: 3.04% Indels: 39

DB: 1 Gaps: 10

US-09-932-678-2 (1-651) x US-08-329-681A-3 (1-2115)

Qy 23 LysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsn----- 39

Db 880 AAAATGGGTAAGGAAAAGACTCACGTTTCGAGCGCCGCGATTAAATTTCCAAACATGGATGCT 939

Qy 40 AspPhePheAsnSerPro---ProArgLysThrVal---ArgPheGlyGlyThrVal--- 56

Db 940 GATTTATATGGGTATAAATGGGCTCGCGATAATGTCCGGCAATCAGGTGCGACAATCTAT 999

Qy 57 -----ThrGluValLeuLeuLysTyrLysLysGlyGluThr 68

Db 1000 CGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAAACATGCGAAAAGGTAGCGTT 1059

Qy 69 AsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIle 88

Db 1060 GCCAAT-----GATGTTACAGATGAGATGGTC 1086

Qy 89 ---IleAsnTrpLeuLeuGluPheArg-----SerSerIleMetTyrLeuThrLysAsp 105

Db 1087 AGACTAAACTGGCTGACGGAATTTATGCTCTTCCGACCATCAAGCATTTTATCCGTACT 1146

Qy 106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrVal 125

Db 1147 CCTGATGATGCATGTTTACTCACCACTGCGATCCCCGGGAAAAACAGCATTCAG---GTA 1203

```
QY 126 ValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145
   ::::::::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1204 TTAGAAGAATATCCTGATTCCAGGTGAAAAATATTGTTGATGCGCTGGCAGTGTTCCTGGC 1263
QY 146 ProCysLeuSerMetIleAlaSerHis-----PheValProProArgValIle----- 161
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1264 CGTTGCAATTCGATTCCCTGTTGTAATGTCCTTTTAACACGATCGCGTATTTTCGTC 1323
QY 162 -----IleLysGluGlyAspValAspValSerAspSerAspAspGlu 175
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1324 GCTCAGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAG 1383
QY 176 AspAspAsnLeuPro 180
   :: | | | |
Db 1384 CGTAATGGCTGGCCT 1398

RESULT 14
US-09-581-909-2
; Sequence 2, Application US/09581909
; Patent No. 6566048
; GENERAL INFORMATION:
; APPLICANT: Graham Keith DIXON
; APPLICANT: John Leslie THAIN
; APPLICANT: John Philip VINCENT
; APPLICANT: No. 6566048bert Friedemann SCHNELL
; APPLICANT: Suberna Jini CHAUDA
; TITLE OF INVENTION: Acetyl-CoA-Carboxylase from Candida Albicans
; FILE REFERENCE: 009901/0270671 - PJF/PHM70303/UST
; CURRENT APPLICATION NUMBER: US/09/581,909
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: PCT/GB98/03857
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: GB 9726897.3
; PRIOR FILING DATE: 1997-12-20
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 2
; LENGTH: 8054
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-581-909-2

Alignment Scores:
Pred. No.: 0.408 Length: 8054
Score: 103.00 Matches: 131
Percent Similarity: 32.99% Conservative: 95
Best Local Similarity: 19.12% Mismatches: 236
Query Match: 3.03% Indels: 224
DB: 4 Gaps: 34

US-09-932-678-2 (1-651) x US-09-581-909-2 (1-8054)
QY 6 LeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerSerAlaValLysLysLeuGly 25
   |||||:::||||| | | | | | | | | | | | | | | | | | | | | |
Db 3705 TTACACTCCAGATTGCCACCTAAATTGGATGACGGATTGACTGCATTG----- 3752
QY 26 AlaSerArgThr-GlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsnSerPr 45
   ||||| | | | | | | | | | | | | | | | | | | | | | |
Db 3753 GTTGAAGAAGAACTCAAGTAGAGGTGCTGAATTCCCTGCTCGTCAAAATTTAAAACTCATC 3812
QY 45 oProArgLysThrValArgPheGlyGlyThrValThrGlu----- 58
   ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 3813 ACCAAATCAATTGCTGAAAAAT-GGTAATGATATATTAGAAGATGTTGTTGCACCATGGT 3871
QY 59 ----ValLeuLeuLysTyrLysLysGly-----GluThrAsnAspPheGluLe 73
   :: | | | | | | | | | | | | | | | | | | | | | | | |
Db 3872 TTCTATTGCCACAAGTTACCAGAATGGTTGTTGTAACACGAATACGATTACTTTGCATC 3931
QY 73 uLeuLysAsnGlnLeuLeuAsp-----ProAspIleLysAspAs 86
   ||| | | | | | | | | | | | | | | | | | | | | | |
Db 3932 TTTGATTAAACGAATATTATGACGTTTGAAAGTTTGTTCAGGTGAAAAATGTTAGAGAAGA 3991
QY 86 pGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAspPh 106
   || | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 3992 TAATGTTATCTTGAATTAAGAGATGAAAAACAAATCT-----GATTT 4033
   | ::::| | | | | | | | | | | | | | | | | | | | | |
QY 106 eGluGlnLeuIleSerIle----- 113
   :::::::| | | | | | | | | | | | | | | | | | | | | |
Db 4034 GAAAAAAGTTATTGCTATTGGTTTGTCTCATTCACGTGTTAGTGCCAAAGAACAATTTTGAT 4093
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 eLeuArgLeu-----ProTrpLeuAsnArgSerGlnThrValValGluG 128
   |||| | | | | | | | | | | | | | | | | | | | | | |
Db 4094 TTTAGCTATTTTGGACATTTATGAACCATTTGTTGCAATCCAACCTCGTCAGTTGCTGCCCTC 4153
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 128 uTyrLeuAlaPheLeuGlyAsnLeu-----Va 137
   |||| | | | | | | | | | | | | | | | | | | | | | |
Db 4154 TATCAGAGAAGCTTTAAAGAACTTGTTCATTAGACCTCGTGTTCGCCAAAGTTGCAATT 4213
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 137 lSerAlaGlnThrValPheLeuArgProCysLeuSerMetIle----- 151
   : | | | | | | | | | | | | | | | | | | | | | | | |
Db 4214 AAAGGCAAGAGAAATTTTAATTCATGTTCTTTTACCTCCATCAAGGAAAGATCCCGATCA 4273
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 -AlaSerHisPheValProProArgValIle---IleLysGluGlyAspValAspValSe 170
   || | | | | | | | | | | | | | | | | | | | | | | |
Db 4274 ATTGGAACATATTTTGAGGTCACTGTTGTTCAAAACCTCTTATGGTGAATTTTTCCTAA 4333
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 rAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190
   : | | | | | | | | | | | | | | | | | | | | | | | |
Db 4334 ACATAGAGAACCATAATTGGAAATATTTCGTGAGGTGTTGATTCCTCCAAACATATTGTTT 4393
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 190 uGlnIleIleAlaArgTyr---ValProSerThrProTrpPheLeuMetProIleLeuVa 209
   : | | | | | | | | | | | | | | | | | | | | | | | |
Db 4394 TGATGTGTTGGCACAAATCTTAATCAATCCAGACCCATGGGTT-----GCCATTGC 4444
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 209 lGluLysPheProPheValArgLysSerGluArgThr-----LeuGluCy 224
   : | | | | | | | | | | | | | | | | | | | | | | | |
Db 4445 TGCCGCTGAAGTTATGTCAGACGTTCATACCGTGTATGATTTGGGTAAATTTGAATA 4504
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 224 sTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLe 244
   : | | | | | | | | | | | | | | | | | | | | | | | |
Db 4505 TCATGTTAATGACAGACTTCCT----- 4526
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 244 uGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGl 264
   ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 4527 -----ATTGTTGAATGGAAATTCAGTTG--GCTAATATGGGAGCCGCTGGTGTA 4576
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 264 uAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGlu----- 281
   ||||| | | | | | | | | | | | | | | | | | | | | |
Db 4577 CGATGCTCAACAGGGTGCT-----GCTGCCGGTGGCGATGATTCGACATCTATGAAACA 4630
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 282 -----GlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLy 297
   || | | | | | | | | | | | | | | | | | | | | | | |
Db 4631 TGCAGCTTCTGTGTCGTGATTGACCTTGTGTTGATTTCTAAACCAGCATTCACACAAG 4690
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 sAlaGly-----ProGluArg---LeuAspGlnMetValHisProValAla----- 311
   : | | | | | | | | | | | | | | | | | | | | | | | |
Db 4691 AACTGGTGTTTTAGTCCAGCAAGACACTTGGATGATGTTGATGAAACTCTTACAGCTGC 4750
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 312 ----- 312
   | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4751 ATTGGAACAATTTCCAACCAGCCGATGCTATTTTCATTTAAAGCAAGGGTGAAACTCAGA 4810
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 312 uArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAs 332
   | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4811 GTTATTAAATGTTTGAATATTGTTCATTACCAGT-----ATTGA 4849
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 332 pGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAs 352
   || | | | | | | | | | | | | | | | | | | | | | | |
Db 4850 TGGTTACTCCGATGAAAAATGAA-----TACTTGAGCAGAGATTAAATGAAATC-- 4895
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 pLysLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCy 372
   || | | | | | | | | | | | | | | | | | | | | | | |
Db 4896 ----- 4900
   | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 372 sSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGlnAs 392
   | : | | | | | | | | | | | | | | | | | | | | | | |
```



Db	4901	CGAATACAAAGAGAGTGTGATTCTGCTGGTGTTCGTCGTGTT-----	4943
QY	392	pProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyr-----	406
Db	4944	-----ACATTTGTTTTTGTCTCATCAAAATGGTCAATATCCTFAAATATTATACTTT	4993
QY	407	-IleGlySerPheLeuAlaArgAlaLysPheIle-----ProLeuIleThrVa	422
Db	4994	TACTGGTCTCTGACTATGAAGAAACAAGGTTATTAGACACATTGAACCAGCTTTGGCTTT	5053
QY	422	lLysSerCysLeuAspLeuLeuValAsnTrp-----	432
Db	5054	CCAAATGGAAATTTGGGAAGATTAGCCAATTTCGATATCAAAACCAATTTTCACTTACAACAG	5113
QY	433	----LeuHisIleTyr-----LeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAs	449
Db	5114	AAACATCCATGATATATGCAATTTGGGAAGATGCTCCTTCTGTATAAAGATT-----	5168
QY	449	pValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValph	469
Db	5169	-----ValValAsnPhe-----PheAlaAlaIleTh	5170
QY	469	eArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLe	489
Db	5171	CACCAGAGGGATTATTAGAACCGGTGTTCTTTAAAGAAGACATTAGCATTAGTGAATATTT	5230
QY	489	uAsnPheGlu-----ArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuPr	507
Db	5231	GATTGCTGAATCCAACAGATTAAATGAATGATATTTTGGATACTTTAGAAAGTATTGACAC	5290
QY	507	oSer-----ValValAsnPhe-----PheAlaAlaIleTh	517
Db	5291	TTCTAAATCTGATTAAACCATATTTTCATTAACTTTTCCCAATGCTTTCAATGTTCAAGC	5350
QY	517	rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLe	537
Db	5351	TTCAGATGTTGAGGCTGCCTTT---GGATCATCTTTAGAAAGATTTGGTAGAAGATTATG	5407
QY	537	uProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAs	557
Db	5408	GAGATT-----AGAGTTACTGGTGTGCTGGAATTAGAAATGTCTGTACTGATCCTCAAG	5461
QY	557	pThrPhePhepro 561	
Db	5462	TACTTCGTGCCCA 5474	
RESULT 15			
US-09-134-001C-335			
; Sequence 335, Application US/09134001C			
; Patent No. 6380370			
; GENERAL INFORMATION:			
; APPLICANT: Lynn Doucette-Stamm et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS			
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: GTC-007			
; CURRENT APPLICATION NUMBER: US/09/134,001C			
; CURRENT FILING DATE: 1998-08-13			
; PRIOR APPLICATION NUMBER: US 60/064,964			
; PRIOR FILING DATE: 1997-11-08			
; PRIOR APPLICATION NUMBER: US 60/055,779			
; PRIOR FILING DATE: 1997-08-14			
; NUMBER OF SEQ ID NOS: 5674			
; SEQ ID NO 335			
; LENGTH: 1692			
; TYPE: DNA			
; ORGANISM: Staphylococcus epidermidis			
US-09-134-001C-335			
Alignment Scores:			
Pred. No.:	0.0323	Length:	1692
Score:	102.50	Matches:	104
Percent Similarity:	35.32%	Conservative:	74
Best Local Similarity:	20.63%	Mismatches:	147

Query Match:	3.01%	Indels:	179
DB:	4	Gaps:	25
US-09-932-678-2 (1-651) x US-09-134-001C-335 (1-1692)			
QY	57	ThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLysAsn	76
Db	415	ACGCAATCTTTACTTAAGCAAAA-----TATCATCTT-----	447
QY	77	GlnLeuLeuAspProAspIleLysAspGlnIleAsnTrpLeuLeuGluPheArg	96
Db	448	CAACTATTAGAT---GATTATGCAGACAATCAGTATTTCAGATTACTTAATCAATATCAA	504
QY	97	SerSerIleMetTyrLeu-----ThrLysAspPheGluGlnLeuIleSer-----	111
Db	505	CTTTCTTATAACCAATATAAAAATAAACGTAAGAATTAGAGGAATTAGAAATCCGCGGAC	564
QY	112	-----IleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValGluGluTyr	129
Db	565	CAAGCTTTATTACAACGATTA-----GACTTA	591
QY	130	LeuAlaPhe---LeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeu	148
Db	592	ATGAAATTTCAATTAGAGGAACCTAACCGAAGCTTCA-----	627
QY	149	SerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAsp	168
Db	628	-----CTGAAAGAAGGCGAAGTGGAC	648
QY	169	ValSerAspSerAsp-----AspGluAspAspAsnLeuProAlaAsnPheAsp	184
Db	649	CAACTTGAATCCGATATTAAAGAAGATTCAAAACTCCGAAAAATTAAATCTAGCTTTAAAC	708
QY	185	ThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeu	204
Db	709	AATGCACATCAAGTTCTA-----ACTGATGAAAAGTGCA	741
QY	205	MetProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys	224
Db	742	ATACCC-----	747
QY	225	TyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeu	244
Db	748	-----GATAGGTTGTACGAATTAAGCAACTACTTTCACAAACGATTAAT-----	789
QY	245	GluLeuIleIleGluLysLeuLeuLysLeu-----AspValAsnAlaSerArgGlnGly	262
Db	790	GATATCGTCCAGAAAAAATTCGTAGATTAAAGAGGACATTAAATCAATTTTACTATTGT	849
QY	263	IleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly	282
Db	850	CTAGAAGATGCAAGCATGAAATT-----	873
QY	283	LeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArg	302
Db	874	-----TACGACGAAATGGCTAACACTGAATTCGAT-----	903
QY	303	LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu	322
Db	904	---GAGCAAGTTTTTAAATGAGTATGAATCCAGATGAATTTACTTAAATAATTTAAACGT	960
QY	323	SerTyrMetLysAspValCys-----TyrVal	331
Db	961	AAATATGGTAAGGATATTACTGAACCTATTGCTTATCAGAGTAAACTTGCAATGAAATT	1020
QY	332	AspGlyLysValAspAsn-----GlyLysThrLysAsp	342
Db	1021	GAT---AAAATAGAAAACTATGAACAAAGTACATCACAAATTAAGGGAAGAAATTAACG	1077
QY	343	LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeu-----	355
Db	1078	CTTTATAACGAAGTGATAGATATAGGAAAAAAACTTCTCAAGAACGTAGCGGTAGCG	1137

QY	356	-----LeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSer	373
Db	1138	AGAGAGTTAAGGGACCATATTGTTTCTGAAATACAA-----	1173
QY	374	PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro	393
Db	1174	---AATTTCACAAATGAAAGATGCTAACCTTGAAATTCGTTTAAACCATTAGATGAACCT	1230
QY	394	Ser-----AsnPro-----	396
Db	1231	ACAATTGAAGGTATTGAATTTGTGGAATTTTAAATTAGTCCAAATCGTGGTGAAACCACTT	1290
QY	397	---AlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLys	415
Db	1291	AAAAGTCTTAATAAAATCGCTTCAGGCGGTGAACCTTCAAGAAATTATGCTAGCTCTAAAA	1350
QY	416	PheIleProLeuIleThrValLysSerCysLeuAspLeu-LeuValAsnTrpLeuHisIle	435
Db	1351	AGTATATTGTTTAAATCACGCGGGCCAAACCGCGATTCTTTTGTGATGAAGTTGACTCGGGT	1410
QY	435	eTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPr	455
Db	1411	GTATCTGGTCAAGCAGCACAAAAAATGGGTGAAAAAATGCGAGATATTGCTCAAT-----	1465
QY	455	oPheTyrSerAlaCysGlnAlaValPheTyrThrPheVal-PheArgHisLysGlnLeuL	475
Db	1466	-ATATACAAAGTCA-----TTTGTATTTCACACTTACCTCAGGTAGCTTCAATGAGTGAC	1518
QY	475	euSerGlyAsnLeuLysGluGlyLeuGln-----TyrLeuGlnSerLeuAsnP	491
Db	1519	CATCATCTTCTTAATAAGCAAGGCATCCAATGCCGATAGAACTACAACCTCAAGTCAAAGAA	1578
QY	491	heGluArg	493
Db	1579	TTGAAAGA	1586



XX (REED/) REEDER R H.  
PA (MOOR/) MOOREFIELD B.  
PA (GREE/) GREENE E A.  
XX Reeder RH, Moorefield B, Greene EA;  
PI  
XX WPI; 2002-681729/73.  
DR P-PSDB; ABB78308.  
XX  
PT Novel RN3 polypeptide, an eukaryotic RNA polymerase I transcription  
PT factor and polynucleotide encoding it useful for diagnosing, treating  
PT hyper and hypoproliferative diseases in mammals and to identify  
PT modulators.  
XX  
PS Claim 6; Page 25-26; 31pp; English.  
XX  
CC The present sequence encodes a human RN3 polypeptide. RN3 is an  
CC eukaryotic RNA polymerase I transcription factor. RN3 is useful for  
CC screening for hypo or hyperproliferative diseases, including cancer,  
CC malignancy, hyperplasia, metaplasia, dysplasia, benign tumour,  
CC hyperproliferative disorder, benign dysproliferative disorder, autoimmune  
CC disease or cardiac disease. It is also used to treat diseases involving  
CC decreased cell proliferation, including degenerative disorders, growth  
CC deficiencies, hypoproliferative disorders, physical trauma, lesions and  
CC wounds. Rn3 polypeptide, nucleic acid or antibody are also useful in  
CC diagnostics, to detect, prognosis, diagnose or monitor various diseases.  
CC RN3 nucleic acids are useful to identify other mammalian genes that  
CC encode RN3-like molecules, to screen for mutations in a RN3 gene that  
CC are associated with certain diseases. The polypeptide is useful as an  
CC immunogen to generate antibodies which immunospecifically bind RN3  
CC polypeptides  
XX  
SQ Sequence 2068 BP; 573 A; 437 C; 485 G; 573 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 0 Length: 2068  
Score: 3403.00 Matches: 651  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-932-678-2 (1-651) x ABV72159 (1-2068)  
  
QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerSerAla 20  
Db 69 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATCGGCGCGTTCGTCTCTGCA 128  
  
QY 21 VallLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
Db 129 GTTAAGAAGCTGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGATGAC 188  
  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
Db 189 TTTTTCATTTCTCCCCCAAGAAAACGTTCGGTTTGGTGGAACTGTGACAGAGTCTTG 248  
  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
Db 249 CTGAAGTACAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGACCAGCTGTTAGAT 308  
  
QY 81 ProAspIleLysAspAspGlnIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
Db 309 CCAGACATAAAGGATGACCAGATCATCAACTGGCTGGCTGTAGAATTCCGTTCTTCTATCATG 368  
  
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleLeuArgLeuProTrpLeuAsn 120  
Db 369 TACTTGCAAAAGACTTTGAGCAACTTATCAGTATTATATTAGATTGCCTTGGTTGAAT 428  
  
QY 121 ArgSerGlnThrValValGluGluThrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
Db 429 AGAAGTCAAAACAGTAGTGAAGAGTATTGGCTTTTCTTGGTAATCTTGTATCAGCACAG 488

QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
Db 489 ACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCCGAGTG 548  
  
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180  
Db 549 ATCAATTAAGGAAGCGCATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCT 608  
  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
Db 609 GCAAATTTTGACACATGTACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 668  
  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
Db 669 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAAATTTCCATTTGTTGAAAAATCAGAGAGA 728  
  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
Db 729 ACACGTGAATGTTACGTTCAATACTTACTAAGGATTAGTGATATATTTCCAACCTTGAGG 788  
  
QY 241 HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
Db 789 CATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTAATGCATCCCGG 848  
  
QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
Db 849 CAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGACAGATTCCACG 908  
  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
Db 909 GAAGGATTGTTTAAATATGGATGAAGATGAAGAACTGAACATGAACAAAGGCTGGTCTCT 968  
  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
Db 969 GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCCTGATGTCTTTG 1028  
  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
Db 1029 GTTTTGTCTACATGAAGGATGTCGTATGTAGATGGTAAGGTTGATAACGGCAAAACA 1088  
  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
Db 1089 AAGGATCTATATCGCGACCTGTATAAACATCTTTTGACAAAACCTCCTGTGTGCCACCCATGCC 1148  
  
QY 361 SerCysHisValGlnPhePheMetPheTyrLeuLeuCysSerPheLysLeuGlyPheAlaGlu 380  
Db 1149 TCCTGCCATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAATTTGGGATTCGCAGAG 1208  
  
QY 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400  
Db 1209 GCATTTTGGAAACATCTCTGGAATAATTCAGGACCCCAAGTAATCCTGCCATCATCAGG 1268  
  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
Db 1269 CAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCTCTTATT 1328  
  
QY 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
Db 1329 ACTGTAAAAATCATGCTTAGATCTTTTGGTTAACTGGCTGCACATATACCTTTAATAACCA 1388  
  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460  
Db 1389 GATTTCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGC 1448  
  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
Db 1449 CAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAAGCAGCTTTTGGCGGGAACCTGAAA 1508  
  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
Db 1509 GAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGTAGAGCCAGCTAAAT 1568  
  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520

Db 1569 CCCCTGAAGATTGGCTGCCCTCAGTGGTTAACTTTTGGTGCATCACAATAAGTAC 1628  
QY 521 GlnLeuValPheCysTyrThrIleGluArgAsnAsnArgGlnMetLeuProValIle 540  
Db 1629 CAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCATT 1688  
QY 541 ArgSerThrAlaGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560  
Db 1689 AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACTTCTTC 1748  
QY 561 PropheAspProCysValLeuLysArgSerLysPheIleAspProIleTyrGlnVal 580  
Db 1749 CCCTTTGATCCCTGTGTGCTGAGAGGCTCAAGAAATTCATTGATCCTATTATCAGGTA 1808  
QY 581 TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
Db 1809 TGGGAAGACATGAGTGTGCTGAGAGCTACAGGAGTTCAAGAAACCCATGAAAGGACATA 1868  
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
Db 1869 GTGGAAGATGAGATGATGATCTTCTGAAAGCGAAGTGCACGAAATGATACCGTGATT 1928  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer 640  
Db 1929 GGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCCTCAAGTAGTGTGGGCTCC 1988  
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
Db 1989 CCACCCGTGTGTACATGCACCCAGTCCCTC 2021

RESULT 2

AAA98384

ID AAA98384 standard; cDNA; 2040 BP.

XX

AC AAA98384;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human RNA polymerase I transcription factor TIF-1A cDNA.

XX

KW RNA polymerase I transcription factor TIF-1A; antitumor; treatment;  
KW antiproliferative; cell proliferation; cancer; tissue regeneration; ss.

XX

OS Homo sapiens.

XX

PN WO200055316-A1.

XX

PD 21-SEP-2000.

XX

PF 08-MAR-2000; 2000WO-DE000767.

XX

PR 17-MAR-1999; 99DE-01011992.

XX

XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

PA Grummt I, Vingron M;

XX

PI WPI; 2000-587527/55.

XX

DR P-PSDB; AAB10936.

XX

PT New DNA encoding the transcription factor TIF-1A, useful for preventing  
PT or treating diseases associated with abnormal cell proliferation,  
PT particularly tumors.

XX

PS Claim 1; Fig 2; 38pp; German.

XX

CC This invention describes a novel DNA sequence (I) that encodes the RNA  
CC polymerase I transcription factor TIF-1A which has antitumor,  
CC antiproliferative and proliferation-inducing activity. The invention also  
CC describes (1) DNA (Ia) encoding a protein (II) with the biological  
CC activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and  
CC able to bind specifically to, and cleave, its transcribed RNA so as to

CC reduce or inhibit synthesis of the corresponding protein; (3) an  
CC antisense RNA (AS) with binding properties similar to R; (4) an  
CC expression vector that contains (I), (Ia) or sequences that encode R or  
CC AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II)  
CC encoded by (I) or (Ia); (7) preparation of TIF-1A or (II) by culturing  
CC cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists  
CC that weaken or block the activity of TIF-1A or (II); (10) a diagnostic  
CC method for detecting abnormal TIF-1A expression; and (11) kit for  
CC carrying out the method in (10). (I), and similar sequences that encode  
CC proteins with equivalent activity, expression vectors containing them, or  
CC their expression products are used to treat or prevent disorders  
CC associated with reduced cellular proliferation, to stimulate cellular  
CC proliferation, and to promote tissue regeneration, e.g. after injury or  
CC radiation therapy. Ribozymes, antisense sequences directed against (I),  
CC also ligands and antagonists of TIF-1A, are used to treat or prevent  
CC disorders associated with excessive cellular proliferation and to inhibit  
CC proliferation, especially in treatment of cancers. (I) and specific  
CC ligands for TIF-1A (particularly antibodies (Ab)) are also useful for  
CC diagnosis of altered TIF-1A expression by (in)direct determination of the  
CC concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for  
CC detecting mutations. Ab can also be used for immunoprecipitation of TIF-  
CC 1A and for isolation of related sequences from cDNA expression libraries.  
CC (I) allows recombinant production of TIF-1A in sufficient quantities for  
CC therapeutic use

SQ Sequence 2040 BP; 563 A; 433 C; 474 G; 568 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.: 0 Length: 2040  
Score: 3391.00 Matches: 649  
Percent Similarity: 99.69% Conservative: 0  
Best Local Similarity: 99.69% Mismatches: 2  
Query Match: 99.65% Indels: 0  
DB: 3 Gaps: 0

US-09-932-678-2 (1-651) x AAA98384 (1-2040)

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20  
Db ATGGCGGCACCGCTGCTTCACACGCGTTGCCGGGAGATGCGCCGCTTCGTCCTCTGCA 99  
QY 21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
Db GTTAAGAAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATGCGTGCAATAGAGATGAC 159  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyThrValThrGluValLeu 60  
Db TTTTTCATTTCTCCCCAAGAAACAACTGTCGTTTGGTGGAACTGTGACAGAGTCTTG 219  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
Db CTGAAGTACAAAAGGGTGAAACAAATGACTTTGAGTTGTGAAGAACCACTGTTAGAT 279  
QY 81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
Db CCAGACATAAAGGATGACCAGATCATCACTGGCTGGCTAGATAATCCGTTCTTATCATG 339  
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn 120  
Db TACTTGACAAAAGACTTTGAGCAACTTATCATGATTTATATTAAGATTGCGTTGGTTGAAT 399  
QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
Db AGAAGTCAAAACAGTAGTGGAGAGATTTGGCTTTTCTTGTTAAATCTTGATATCAGCACAG 459  
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
Db ACTGTTTCTCTCAGACCGGTGCTCAGCATGATTTGCTTCCCATTTTTGTCCTCCCCGAGTG 519  
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180  
Db ATCATTAAAGGAAGCGGATGTAGATGTTTCAGATTCTGTATGATGAGATGATAATCTTCCT 579

QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleLeAlaArgTyrValProSerThr 200  
Db 580 GCAAAATTTGACACATGTACAGAGCCTTGCAATAATAGCAAGATATGTACCATCGACA 639  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
Db 640 CCGTGGTTTCTCATGCCAATACTGGTGAAAAAATTTCCATTTGTCGAAAAATCAGAGAGA 699  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
Db 700 ACACCTGGAATGTTACGTTTCATNACTTACTAAGGATTAGTGATATATTTTCCAAACCTTGAGG 759  
QY 241 HisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
Db 760 CATGAAATTTCTGGAGCTTATATTGAAAAAATACTCAAGTTGGATGTGAATGCATCCCGG 819  
QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
Db 820 CAGGGTATTGAAGATGCTGAAGAAACAGCAAACTCAAACCTTGTGTGGGACAGATTCCACG 879  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
Db 880 GAAGGATTGTTTAAATATGGATGAAGATGAAGAAAACTGAACATGAACAAAGGCTGGTCTCT 939  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
Db 940 GAACGGCTCGACACAGATGGTGATCCTCTGTAGCCGAGCGCTGGACATCCTGTATGTTCTTTG 999  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
Db 1000 GTTTTGTCTTACATGAAGGATGTCTGTCTATGTAGATGGTAAGTTGTATAACGGCAAAACA 1059  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThrHisAla 360  
Db 1060 AAGGATCATATATCGGACCTGATAAAACATCTTTGACAAACTCCTGTTGCCACCCATGCC 1119  
QY 361 SerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
Db 1120 TCCTGCCATGTACAGTTTTCATGTTTAACTCTGTAGTTTCAAATTTGGGATTCGCAGAG 1179  
QY 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400  
Db 1180 GCATTTTGGAAACATCTCTGAAAAAATTTGCAGGACCCCAAGTAATCCTGCCATCATCAGG 1239  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
Db 1240 CAGGCTGCTGGAAAATTATATTGGAAGCTTTTGGCAAGAGTAAATTTATCTCTTATT 1299  
QY 421 ThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
Db 1300 ACTGTAAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAG 1359  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPropheTyrSerAlaCys 460  
Db 1360 GATTTCGGGAACAAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC 1419  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
Db 1420 CAAGCTGTGTTCTACNCCTTTGTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAA 1479  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
Db 1480 GAAGGTTTGAGATATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCAGCTAAAT 1539  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr 520  
Db 1540 CCCCTGAAGATTGCTTGCCTGCCTCAGTGGTTAACTTTTTTGTGCAATCACAAATAAGTAC 1599  
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540  
Db 1600 CAGCTCGTCTTCTGTCTACCATCATTTAGAGGAACAATCGCCAGATGCTGCCAGTCAAT 1659  
QY 541 ArgSerThrAlaGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560

Db 1660 AGGAGTACCGCTGGAGGAGACTCAGTCAGATCTGCACAAACCCGCTGGACACCTTCTTC 1719  
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
Db 1720 CCTTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTG 1779  
QY 581 TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
Db 1780 TGGGAACACATGAGTGTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 1839  
QY 601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
Db 1840 GTGGAAGATGAAGATGATGACTTTCTGAAAGGGCAAGTGCCCCAGAAATGATACCGTGATT 1899  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer 640  
Db 1900 GGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTGTGGGCTCC 1959  
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
Db 1960 CCACCCGTGTTGTACATGCAACCCAGTCCCTC 1992  
RESULT 3  
ABS78724  
ID ABS78724 standard; DNA; 1770 BP.  
XX  
AC ABS78724;  
XX  
DT 16-DEC-2002 (first entry)  
XX  
DE DNA encoding human NOVX3 protein.  
XX  
KW Human; NOVX; human disease; NOVX-associated disorder; cancer; addiction;  
KW Hodgkin disease; Von Hippel-Lindau syndrome; Alzheimer's disease; stroke;  
KW tuberos sclerosis; hypercalcaemia; Parkinson's disease; depression;  
KW Huntington's disease; cerebral palsy; epilepsy; Iesch-Nyhan syndrome;  
KW multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety; pain;  
KW obesity; Crohn's disease; osteoporosis; inflammatory bowel disease;  
KW infertility; inflammatory bowel disease; atherosclerosis; hypertension;  
KW scleroderma; haemophilia; diabetes; pancreatitis; autoimmune disease;  
KW asthma; arthritis; immunodeficiency; HIV; viral infection; neurogenesis;  
KW bacterial infection; parasitic infection; graft-versus-host disease;  
KW cell differentiation; cell proliferation; haematopoiesis; wound healing;  
KW angiogenesis; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200272770-A2.  
XX  
PD 19-SEP-2002.  
XX  
PF 08-MAR-2002; 2002WO-US0007283.  
XX  
PR 08-MAR-2001; 2001US-0274281P.  
PR 09-MAR-2001; 2001US-0274849P.  
PR 12-MAR-2001; 2001US-0275235P.  
PR 13-MAR-2001; 2001US-0275579P.  
PR 13-MAR-2001; 2001US-0275601P.  
PR 14-MAR-2001; 2001US-0276000P.  
PR 20-MAR-2001; 2001US-0277239P.  
PR 20-MAR-2001; 2001US-0277327P.  
PR 20-MAR-2001; 2001US-0277338P.  
PR 21-MAR-2001; 2001US-0277791P.  
PR 22-MAR-2001; 2001US-0277833P.  
PR 23-MAR-2001; 2001US-0278152P.  
PR 26-MAR-2001; 2001US-0278894P.  
PR 27-MAR-2001; 2001US-0279036P.  
PR 28-MAR-2001; 2001US-0279344P.  
PR 30-MAR-2001; 2001US-0280233P.  
PR 02-APR-2001; 2001US-0280802P.  
PR 02-MAY-2001; 2001US-0288148P.  
PR 31-MAY-2001; 2001US-0294821P.



PR 31-OCT-2001; 2001US-0335302P.  
PR 04-DEC-2001; 2001US-0338375P.  
PR 07-MAR-2002; 2002US-00094466.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
XX Spytek KA, Vernet CA, Tchernev VT, Malyankar UM, Gerlach VL;  
PI Li L, Zerhusen BD, Patturajan M, Gusev VY, Kekuda R, Pena CEA;  
PI Zhong M, Gangolli EA, Taupier RJ;  
XX  
XX WPI; 2002-713508/77.  
DR P-PSDB; ABG97480.  
DR  
XX  
PT New NOVX polypeptides and polynucleotides, useful for preventing,  
PT diagnosing or treating NOVX-associated disorders, e.g. diabetes, multiple  
PT sclerosis, atherosclerosis, cancer, infections, osteoporosis or  
PT Parkinson's disease.  
XX  
PS Claim 22; Page 98; 266pp; English.  
XX  
CC The present invention relates to a new polypeptide (NOVX). The NOVX  
CC polypeptide, nucleic acid and antibody are useful in the manufacture of a  
CC medicament for treating a syndrome associated with a human disease,  
CC preferably a NOVX-associated disorder. The NOVX nucleic acids,  
CC polypeptides and antibodies are useful for treating, preventing or  
CC diagnosing diseases such as cancers, Hodgkin disease, Von Hippel-Lindau  
CC syndrome, Alzheimer's disease, stroke, tuberosus sclerosis,  
CC hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral  
CC palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-  
CC telangiectasia, leukodystrophies, addiction, anxiety, depression, pain,  
CC obesity, Crohn's disease, osteoporosis, inflammatory bowel disease,  
CC infertility, inflammatory bowel disease, atherosclerosis, hypertension,  
CC scleroderma, haemophilia, diabetes, pancreatitis, autoimmune disease,  
CC asthma, arthritis, immunodeficiencies, HIV, viral, bacterial or parasitic  
CC infections, or graft-versus-host disease. The nucleic acids and  
CC polypeptides may also be used as targets for the identification of small  
CC molecules that modulate or inhibit e.g. neurogenesis, cell  
CC differentiation, cell proliferation, haematopoiesis, wound healing and  
CC angiogenesis, in gene therapy, in generation of antibodies that bind  
CC immunospecifically to NOVX substances for use in therapeutic or  
CC diagnostic methods. The nucleic acids are further used as hybridisation  
CC probes, in chromosome mapping, tissue typing, preventive medicine, and  
CC pharmacogenomics. The present nucleic acid sequence encodes a human NOVX  
CC protein of the invention  
XX  
SQ Sequence 1770 BP; 492 A; 373 C; 413 G; 492 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2.43e-296 Length: 1770  
Score: 2839.00 Matches: 556  
Percent Similarity: 85.41% Conservative: 0  
Best Local Similarity: 85.41% Mismatches: 5  
Query Match: 83.43% Indels: 90  
DB: 6 Gaps: 1

US-09-932-678-2 (1-651) x ABS78724 (1-1770)

QY 1 MetAlaLaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20  
Db 40 ATGGCGGACCGCTGCTTCACACGCGTTTGC CGGAGATGCGGCCGCTTCCTCTGCA 99  
QY 21 VallLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
Db 100 GTTAAGAGCTGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGTGAGAAATGAC 159  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
Db 160 TTTTTCATCTCCCCAAGAAAAAAGCTGTTGCGTTTGGTGGAACTGTGACAGAGTCTTG 219  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
Db 220 CTGAAGTACAAAAAGGGTGAAACAAATGACTTGGATTGTTGAAGAACCGAGCTGTTAGAT 279

QY 81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
Db 280 CCAGACATAAAGGATGACCAGATCATCACTGGCTGCTAGAAATCCGTTCTTCTATCATG 339  
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn 120  
Db 340 TACTTTGACAAAAAGACTTTTGAGCAACTTATCAGTATTATATTAAAGATTGCCTTGGTTGAAT 399  
QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
Db 400 AGAAGTCAAACAGTAGTGGAAAGAGTATTTCGCTTTTCTTGGTAATCTTGTATCAGCACAG 459  
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
Db 460 ACTGTTTTCTCAGACCCGCTGCTCAGCAATGATTCCTTCCCATTTTGTGCCCTCCCCGAGTG 519  
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180  
Db 520 ATCATTAAGGAAGCGGATGTAGATGTTTTCAGATTCTGATGATGAAGATGATAATCTTCCT 579  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
Db 580 GCAAAATTTTGACACATGTCAGAGAGCCCTTGCAAAATAATAGCAAGATATGTACCATCGACA 639  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
Db 640 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAAATTTCCATTGTTTCAAAATCAGAGAGA 699  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
Db 700 ACACCTGGAATGTTACGTTCAATAACTTACTAAAGATTAGTGTATATTTTCCAACTTGAGG 759  
QY 241 HisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
Db 760 CATGAAATCTGGAGCTTATTATTGAAAAAATACTCAAGCTGGATGTGAATGCATCCCGG 819  
QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
Db 820 CAGGGTATTGAAGATGCTCAAGAAACAGCAAAATCAAACCTTGTGGTGGGACAGATTCCACG 879  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
Db 880 GAAGGATTGTTTAATATG----- 897  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
Db 897 ----- 897  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
Db 897 ----- 897  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThrHisAla 360  
Db 897 ----- 897  
QY 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
Db 898 -----GGATTCGCAGAG 909  
QY 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400  
Db 910 GCATTTTGGAAACATCTTTGGAAAACTTGCAGGATCCAAGTAATCCTGCCATCATCAGG 969  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
Db 970 CAGGCTGCTGGAAATATATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTTCTCTTATT 1029  
QY 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
Db 1030 ACTGTAACCAACCATGCCTAGATCTTTTGGTTAACTGGTGACATATACCTTAATAACACAG 1089  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460

Db 1090 GATTCGGGAACAAAGGCATTCTCGGATGTTGCTCTCCATGGACCAATTACTCAGCCTGC 1149  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
Db 1150 CAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGCGGAACCTGAAA 1209  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
Db 1210 GAAGGTTTGAGTATCCTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1269  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr 520  
Db 1270 CCCCTGAAGATTTCCTGCCCCCTCAGTGGTTAACTTTTGTGCAATCACAAATAAGTAC 1329  
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540  
Db 1330 CAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCAAT 1389  
QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560  
Db 1390 AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTC 1449  
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
Db 1450 CCCTTTGATCCCTGTGTGCTGAAGAGTCAAGAGTCAAGAAATTCATTGATCCTATTATCAGGTG 1509  
QY 581 TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspile 600  
Db 1510 TGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAGAAACCCATGAAAGGACATA 1569  
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
Db 1570 GTGGAAGATGAAGATGATGACTTTCTGAAAGCGGAAGTGCCTCAAGTAGTGTTGGGCTCC 1629  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer 640  
Db 1630 GGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCTTCAAGTAGTGTTGGGCTCC 1689  
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
Db 1690 CCACCCGTGTTGTACATGCAACCCAGTCCCTC 1722

RESULT 4

AAC76937  
ID AAC76937 standard; cDNA; 1418 BP.

AC AAC76937;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4983.

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.

XX Homo sapiens.

XX WO200058473-A2.

PN 05-OCT-2000.

PD

XX 31-MAR-2000; 2000WO-US008621.  
PF  
XX 31-MAR-1999; 99US-0127607P.  
PR 02-APR-1999; 99US-0127636P.  
PR 05-APR-1999; 99US-0127728P.  
PR 30-MAR-2000; 2000US-00540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach M;  
XX  
DR WPI; 2000-602362/57.  
DR P-PSDB; AAB42728.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.

PS Claim 5; Page 4157-4158; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX SQ Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.81e-223 Length: 1418  
Score: 2162.00 Matches: 413  
Percent Similarity: 98.34% Conservative: 2  
Best Local Similarity: 97.87% Mismatches: 6  
Query Match: 63.53% Indels: 1  
DB: 3 Gaps: 0

US-09-932-678-2 (1-651) x AAC76937 (1-1418)

Qy 202 TrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArgThr 221  
Db 3 TGGTTTCTCATGCCCACTACTGGTGGAAAAATTTCCATTGTTTCGAAAAATCAGAGAAACA 62  
Qy 222 LeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThr-LeuArgHi 241  
Db 63 CTGGAATGTTACGTTTCATAACTTACTAAGGATTAGTGTATATTTTCCAACCTTTGAGGCA 122  
Qy 241 sGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgI 261  
Db 123 TGAATTTCTGGAGCTTATTATTGAAAAAATACTCAAGTTGGATGTGAATGCCATCCCGCA 182  
Qy 261 nGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGl 281  
Db 183 GGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTTTGGTGGGACAGATTCACCGGA 242  
Qy 281 uGlyLeuPheAsnMetZaspGluAspGluGluThrGluHisGluThrLysAlaGlyProGl 301  
Db 243 AGGATTGTTTATATATGATGATGAAGATGAAGAAACCTGAACATGAACAAAGGCTGGTCTCTGA 302

QY 301 uArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuVa 321  
Db ACGGCTCGACCAAGATGGTGCACTCTGTAGCCGAGCGCTGGACATCCTGATGICTTTGGT 362  
QY 321 lLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLy 341  
Db TTTGTCTACATGAAGGATGTCTGCTATGTAGTGTGAAGTTGATAACGGCAAAACAAA 422  
QY 341 sAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSe 361  
Db GGATCTATATCGGACCTGTATAACATCTTTTGACAAACTCCTGTTGCCCCACCCATGCCTC 482  
QY 361 rCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAl 381  
Db CTGCCATGTACAGTTTTTTCATGTTTACCTCTGTAGTTTCAAANTGGGATTCGAGAGGC 542  
QY 381 aPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGl 401  
Db ATTTTGGAAACATCTCTGGAAAAAATTGCAGGACCCAAAGTAATCTCTGCCATCATCAGGCA 602  
QY 401 nAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleTh 421  
Db GGCTGCTGGAATTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATCTCTTATTATAC 662  
QY 421 rValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnGlnAs 441  
Db TGTAAATCATCGCTAGATCTTTGGTTAACTGGCTGCACATATACCTTAATAACACAGGA 722  
QY 441 pSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGl 461  
Db TTCGGGAACAAAGGCATTCTGCGATGTTGCTCCATGGACCATTATTACTAGCCTGCCA 782  
QY 461 nAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGl 481  
Db AGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGA 842  
QY 481 uGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPr 501  
Db AGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCC 902  
QY 501 oLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGl 521  
Db CCTGAAGATTGCCTGCCCTCAGTGGTTAACTTTTGTCTGCAATCACAATAAGTACCA 962  
QY 521 nLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleAr 541  
Db GCTCGTCTTCGTCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCATTAG 1022  
QY 541 gSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePr 561  
Db GAGTACCGCTGGAGGAGACTCAGTGCAGACTGCACAAACCCACTGGACACCTTCTTCCC 1082  
QY 561 oPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTr 581  
Db CTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATTATCAGGTGTG 1142  
QY 581 pGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleVa 601  
Db GGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAGAAACCCCATGAAAAAGGACATAGT 1202  
QY 601 lGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGl 621  
Db GGAAGATGAAGATGATGACTTTCTGAAAGCGGAAATTCGCCAGAAATTAGTAGTAGTGG 1262

RESULT 5  
ADD29697  
ID ADD29697 standard; mRNA; 1582 BP.  
XX

AC ADD29697;  
XX 15-JAN-2004 (first entry)  
DT  
XX Human tumour suppressor mRNA SEQ ID NO:153.  
DE ss; human; tumour suppressor; cancer; cytostatic; gene therapy.  
XX Homo sapiens.  
OS WO2003058201-A2.  
PN 17-JUL-2003.  
XX 31-DEC-2002; 2002WO-US041825.  
PF 31-DEC-2001; 2001US-0345317P.  
XX (QUAR-) QUARK BIOTECH INC.  
PA (CLEV-) CLEVELAND CLINIC FOUND.  
XX Feinstein E, Gudkov AV;  
PI WPI; 2003-598393/56.  
XX  
PT Diagnosing cancer comprises determining the polypeptide or polynucleotide levels e.g., hepatic lipase, in a sample from a subject, where a higher level compared to that in a subject free of cancer is indicative of cancer.  
PT  
XX Disclosure; SEQ ID NO 153; 272pp; English.  
PS  
XX The invention relates to a novel method for diagnosing a cancer in a subject. the method comprises determining, in a sample from the subject, the level of at least one polypeptide, where a higher level of the polypeptide compared to the level of the polypeptide in a subject free of cancer is indicative of cancer. The polypeptide is selected from any of the polypeptides encoded by the polynucleotides listed in the specification and polypeptides which are at least 70% homologous to the polypeptides. The method of the invention has cytostatic activity, and may have a use in gene therapy. The method is useful in identifying markers specific for one or several types of cancer, depending on the tissue origin, which may be used in numerous diagnostic and prognostic applications as well as cancer type-specific targets for therapeutic intervention. The compounds that modulate the activity of a tumour suppressor gene are useful in the treatment of cancer or as anti-cancer drugs. The present sequence represents a polynucleotide of the invention.  
XX  
SQ Sequence 1582 BP; 364 A; 253 C; 296 G; 389 T; 0 U; 280 Other;  
  
Alignment Scores:  
Pred. No.: 8.56e-206 Length: 1582  
Score: 2001.00 Matches: 410  
Percent Similarity: 79.34% Conservative: 1  
Best Local Similarity: 79.15% Mismatches: 106  
Query Match: 58.80% Indels: 1  
DB: 9 Gaps: 0  
  
US-09-932-678-2 (1-651) x ADD29697 (1-1582)  
  
QY 1 MetAlaAlaProLeuLeuHisThrArg-LeuProGlyAspAlaAlaSerSerSerAl 20  
Db ATGGCGGACCGCTGCTTCACACGTTGTTGTGCGGAGATGCGCGCTTCGTCCTCTGC 82  
QY 20 aValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAs 40  
Db AGTCAAGACGCTGGCGCGTCGAGGACTGGGATTTCAATATGCGTGCATTAGAGATGA 142  
QY 40 pPhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValle 60  
Db TTTTTTCAATTCTCCCCCAAGAAAAAAGTGTTCGGTTGGTGGAACTGTGACAGAGTCTT 202  
QY 60 uLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAs 80

Db 203 GCTGAAGTACAAAAAGGGTGAACAAATGACTTTGAGTGTGTTGAAGAACCCAGCTGTTAGA 262

Qy 80 pProAspIleLysAspAspGlnIleAAsnTrpLeuLeuGluPheArgSerSerIleMe 100

Db 263 TCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAAATTCCTGTCAT 322

Qy 100 tTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAs 120

Db 323 GTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTGAGATTGCTTGGTTGAA 382

Qy 120 nArgSerGlnThrValValGluGluThrLeuAlaPheLeuGlyAsnLeuValSerAlaGl 140

Db 383 TAGAAGTCAAAACAGTAGTGAAGAGTATTGGCTTTTCTGGTAATCTGTATCAGCATA 442

Qy 140 nThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVa 160

Db 443 GACTGTTTCTTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCTGAGT 502

Qy 160 lIleIleLysGluGlyAspValSerAspSerAspGluAspAspAsnLeuPr 180

Db 503 GATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCC 562

Qy 180 oAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerTh 200

Db 563 TGCAAAATTTGACACATATCACAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGC 622

Qy 200 rProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluAr 220

Db 623 ACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTTCGAAAAATCAGAG 682

Qy 220 gThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuAr 240

Db 683 AACACTGGAATGTTACGTTCAATACTTACTAAGGATTAGTGATATATTTTCCAACCTTG 742

Qy 240 gHisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerAr 260

Db 743 GCATGAAATCTGGAGCTTATTATTGAAAACTACTCAAGCTGGATGTGAATGCATCCCG 802

Qy 260 gGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerTh 280

Db 803 GCAGGGTATTGAAGATGCTGAAGAAACAGCAAACTCAACTGTGTGGGACAGATTCCAC 862

Qy 280 rGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPr 300

Db 863 GGAAGGATTGTTTAAATATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 922

Qy 300 oGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerJe 320

Db 923 NNN 982

Qy 320 uValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysTh 340

Db 983 NNN 1042

Qy 340 rLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAl 360

Db 1043 NNN 1102

Qy 360 aSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGl 380

Db 1103 NNN 1162

Qy 380 uAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleAr 400

Db 1163 GGCATTTTGGAAACATCTTTGGAAAACTTCAGGATCCCAAGTAACTCTGCGCATCATCAG 1222

Qy 400 gGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIl 420

Db 1223 GCAGGCTGCTGGAAATTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTTCTTTAT 1282

Qy 420 eThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGl 440

Db 1283 TACTGTAAACCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 1342

Qy 440 nAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCy 460

Db 1343 GGATTCGGGAACAAGGCATTCTCGGATGTTGCTCTCCATGGACCATTTTACTCAGCCTG 1402

Qy 460 sGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLy 480

Db 1403 CCAAGCTGTGTTCTACACCTTTGTTTTAGACACACAGCAGCTTTTGGCGGAAACCTGAA 1462

Qy 480 sGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAs 500

Db 1463 AGAAGGTTTGAGTATCCTCAGAGTCTGAATTTTGGCGGATAGTGATGAGCCAGCTAAA 1522

Qy 500 nProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThr 517

Db 1523 TCCCTGGAAGATTGCTGCTGCCCTCAGTGGTTAACTTTTGTGCTGCAATCACA 1574

RESULT 6

AAS01562

ID AAS01562 standard; cDNA; 1423 BP.

XX

AC AAS01562;

XX

DT 18-JUL-2001 (first entry)

XX

DE Human secretory molecule cDNA sptm #52.

XX

KW Human; secretory molecule; sptm; SPTM; library screening; gene therapy;

KW cell signalling; cell proliferative disorder; atherosclerosis; cancer;

KW immune system disorder; AIDS; neurological disorder; Alzheimer's disease;

KW nervous system disease; mental retardation; developmental disorder;

KW neuromuscular disorder; microarray; Incyte ID number 4287452dec; ss.

XX

OS Homo sapiens.

XX

PN WO200123558-A2.

XX

PD 05-APR-2001.

XX

PF 19-SEP-2000; 2000WO-US025610.

XX

PR 28-SEP-1999; 99US-0156624P.

PR 28-SEP-1999; 99US-0156625P.

PR 02-DEC-1999; 99US-0168611P.

PR 02-DEC-1999; 99US-0168613P.

PR 02-DEC-1999; 99US-0168614P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;

PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;

PI Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR, Roseberry AM;

PI Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK, Amshey S;

PI Fong WT;

XX

DR WPI; 2001-258134/26.

XX

PT New secretory polynucleotides (SPTM) and the polypeptides they encode,

PT for use in inducing antibodies and screening libraries of compounds.

XX

PS Claim 1; Page 152; 161pp; English.

XX

CC The present sequence for human secretory molecule cDNA sptm #52 (Incyte

CC ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences (AAS01511-

CC AAS01573) which encode for the secretory polypeptides SPTM. The sptm

CC polynucleotides are useful for screening a compound for effectiveness in

CC altering expression of a target polynucleotide, where the target

CC polynucleotide comprises sptm. Sptm is also useful in a method for

CC assessing the toxicity of a test compound. Sptm and its fragments or

CC complementary sequences, may be used to identify the presence of and/or

CC determine the degree of similarity between two nucleic acid sequences.

CC Sptm can also be used for a variety of diagnostic and therapeutic

CC purposes, e.g. diagnosing a particular condition, disease or disorder  
CC associated with cell signalling, such conditions include cell  
CC proliferative disorders such as atherosclerosis, and cancers including  
CC leukaemia, an immune system disorder e.g. acquired immunodeficiency  
CC syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's  
CC disease, nutritional and metabolic disease of the nervous system, mental  
CC retardation and other developmental disorders, and muscular dystrophy and  
CC other neuromuscular disorders. Sptm can also be used to design probes  
CC useful in diagnostic assays, which may be used to monitor the progress of  
CC conditions or disorders associated with abnormal levels of expression of  
CC sptm. In addition sptm encoding SPTM may be used for somatic or germline  
CC gene therapy, for inducing antibodies, or in microarrays  
XX  
SQ Sequence 1423 BP; 417 A; 247 C; 320 G; 439 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,76e-203 Length: 1423  
Score: 1973.00 Matches: 393  
Percent Similarity: 80.53% Conservative: 0  
Best Local Similarity: 80.53% Mismatches: 4  
Query Match: 57.98% Indels: 91  
DB: 4 Gaps: 1

US-09-932-678-2 (1-651) x AAS01562 (1-1423)

QY 31 IleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50  
Db 194 ATTTCAAATATCGTGCATTAGAGATGATTTTTCATCTCCCCAAGAAAACTGTT 253  
QY 51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAsp 70  
Db 254 CGGTTTGGTGAACGTGTGACAGAGTCTGTGAGTACAAAAAGGGTGAACAATGAC 313  
QY 71 PheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleAsn 90  
Db 314 TTTGAGTGTGTAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAAC 373  
QY 91 TrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIle 110  
Db 374 TGGCTGCTAGAAATCCGTTCTTCTATCATGTACTTGACAAAAGACTTTTGAGCAACTTATC 433  
QY 111 SerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeu 130  
Db 434 AGTATTATATTAAAGATTGCGCTTGGTTGAAGTAGAAGTCAAACAGTAGTGAAGAGTATTTG 493  
QY 131 AlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150  
Db 494 GCTTTTCTGGTAATCTGTATCAGCACAGACTGTTTTCCTCAGACCCGTCTCAGCATG 553  
QY 151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValSer 170  
Db 554 ATTGCTCCCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGGATGATGTTTCA 613  
QY 171 AspSerAspAspGluAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeu 190  
Db 614 GATTCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTACAGAGCCCTTG 673  
QY 191 GlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGlu 210  
Db 674 CAAATAATAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCATACTGGTGAA 733  
QY 211 LysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeu 230  
Db 734 AAATTTCCATTTGTTCCGAAAATCAGAGAGAACACTGGATGTTACGTTACATAACTTACTA 793  
QY 231 ArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleGluLys 250  
Db 794 AGGATTAGTGATATTTTCCAACCTTGAGCATGAAATCTGGAGCTTATTATTGAAAAA 853  
QY 251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAla 270  
Db 854 CTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCA 913

QY 271 ThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGlu 290  
Db 914 AATCAAACTTGTGGTGGACAGATTCCACGGAAGGATGTTTAATATG----- 961  
QY 291 GluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProVal 310  
Db 961 ----- 961  
QY 311 AlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330  
Db 961 ----- 961  
QY 331 ValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle 350  
Db 961 ----- 961  
QY 351 PheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyr 370  
Db 961 ----- 961  
QY 371 LeuCysSerPheLysLeuGlyPheAlaGluAla-PheLeuGluHisLeuTrpLysLysLe 390  
Db 962 -----GGATTCGACAGAGCCATTTTGGAAACATCTTTGGAAAAACTT 1003  
QY 390 uGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPh 410  
Db 1004 GCAGGATCCAAAGTAATCTGCCATCATCAGGAGGCTGTGGAAATATATTTGGAAGCTT 1063  
QY 410 eLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuVa 430  
Db 1064 TTTGGCAAGAGCTAAATTTATTTCTCTTATTACTGTAAAAACCATGCCTAGATCTTTTGGT 1123  
QY 430 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVa 450  
Db 1124 TAACTGGCTGCACATATACCTTAATAACCCAGGATTCGGGAACAAAGGCATTTCTGCGATGT 1183  
QY 450 lAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheAr 470  
Db 1184 TGCTCTCCATGGACCATTTTACTACGCTGCCAGCTGTGTTCTACACCTTTGTTTTTAG 1243  
QY 470 gHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAs 490  
Db 1244 ACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAA 1303  
QY 490 nPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValVa 510  
Db 1304 TTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGT 1363  
QY 510 lAsnPhePheAlaAlaIleThr 517  
Db 1364 TAACTTTTGTGTCATCAATCAC 1385

RESULT 7

AAS44584  
ID AAS44584 standard; DNA; 1498 BP.  
XX  
AC AAS44584;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human full-length polynucleotide sequence #9.  
XX  
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;  
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;  
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
KW cytostatic; antirheumatic; antiarthritic; vulnery; antinflammatory;  
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;  
KW immunostimulant; analgesic; gene therapy.



OS Homo sapiens.  
XX PN WO200164834-A2.  
XX PD 07-SEP-2001.  
XX PF 26-FEB-2001; 2001WO-US004926.  
XX PR 28-FEB-2000; 2000US-00515126.  
XX PR 18-MAY-2000; 2000US-00577409.  
XX PR 17-JUN-2000; 2000US-00597707.  
XX PR 14-JUL-2000; 2000US-00616807.  
XX PR 19-SEP-2000; 2000US-00664641.  
XX (HYSE-) HYSEQ INC.  
XX PA  
XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI PI Xue AJ, yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
PI PI Drmanac R;  
XX DR WPI; 2001-589862/66.  
DR DR P-PSDB; AAU27684.  
XX PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
PT prepared from various human tissues, for diagnosis, treatment of cancer,  
PT neurological, inflammatory disorders and for use in arrays for detection.  
XX  
PS Claim 1; SEQ ID NO 9; 153pp; English.  
XX  
CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and  
CC contig polynucleotides encoding polypeptides of the invention. The DNA  
CC and protein sequences are useful for the treatment, diagnosis and  
CC prevention of various types of disorder in a mammalian subject such as a  
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers  
CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such  
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,  
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system  
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's  
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and  
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's  
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory  
CC bowel disease. The sequences exhibit activity relating to angiogenesis,  
CC cell proliferation, cell differentiation, stem cell growth factor,  
CC activin or inhibin. Therefore, they can be used to manipulate stem cells  
CC in culture to give rise to neuroepithelial cells that can be used to  
CC augment or replace cells damaged by illness, accidental damage or genetic  
CC disorders. The sequences may also be used for regeneration of bone,  
CC cartilage, tendons and ligaments and in tissue repair and burn healing.  
CC Note: Some sequences for this patent did not form part of the printed  
CC specification, but were obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences  
SQ Sequence 1498 BP; 417 A; 332 C; 325 G; 424 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 4.3e-180 Length: 1498  
Score: 1763.00 Matches: 339  
Percent Similarity: 97.98% Conservative: 1  
Best Local Similarity: 97.69% Mismatches: 7  
Query Match: 51.81% Indels: 1  
DB: 4 Gaps: 0  
  
US-09-932-678-2 (1-651) x AAS44584 (1-1498)  
  
QY 223 GluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGlu 242  
DB 4 GAAAGTTACGTTTCATACTTACAGAGGATTAGTGATATATTTTCCACCTG-AGGCATGAA 62  
  
QY 243 IleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGly 262  
DB 63 ATTCTGAGGCTTATATGGAAAACTACTCAAGTTGGATGTGATGCATCCCGCAGGGT 122  
  
QY 263 IleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly 282

Db 123 ATTGAAGATGCTGAAGAAACAGCAACTCAAACCTTGTTGGTGACAGATTCCACGGAAGGA 182  
QY LeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArg 302  
Db 183 TTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAACAAAGGCTGCTCTGAACGG 242  
QY LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322  
Db 243 CTCGACCAGATGGTGATCCTGTAGCCGAGCGCTGGACATCCTGATGTCTTTGGTTTGG 302  
QY SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp 342  
Db 303 TCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGCAAAACAAAGGAT 362  
QY LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCys 362  
Db 363 CTATATCGCGACCTGATAAACAATCTTTGACAAACTCCTGTTGCCACCCATGCCTCCTGC 422  
QY HisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPhe 382  
Db 423 CATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAATTTGGGATTCGCAGAGGCAATT 482  
QY LeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAla 402  
Db 483 TTGGAACATCTCTGGAATAAATTTGAGGACCCCAAGTAATCCTGCCATCATCAGGCGAGCT 542  
QY AlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal 422  
Db 543 GCTGGAATATATATTGGAAGCTTTTGGCAAGAGAGTAATAATTTATCTCTTATTACTGTA 602  
QY LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSer 442  
Db 603 AAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCG 662  
QY GlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462  
Db 663 GGAACAAAGGCAATCTGCGATGTTGCTCTCCATGGACCACTTTTACTCAGCCTGCCAGCT 722  
QY ValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482  
Db 723 GTGTTCTACACCTTTGTTTTTAGACACACAGCAGCTTTTGAGCGGAAACCTGAAAGAGGT 782  
QY LeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeu 502  
Db 783 TTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGTAGTACCCAGCTAAATCCCTG 842  
QY LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeu 522  
Db 843 AAGATTTGCCCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAATAAGTACCAGCTC 902  
QY ValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSer 542  
Db 903 GTCTTCTGCTACACCATCATCGAGAGGAACAATCGCCAGATGCTGCCAGTCAATAGGAGT 962  
QY ThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPhe 562  
Db 963 ACCACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCACTGGACACCTTCGTCCTCCCTTT 1022  
QY AspProCysValLeuLysArg 569  
Db 1023 GATCCCTGTGTGCTGAAGAGA 1043  
  
RESULT 8  
AAS87641  
ID AAS87641 standard; cDNA; 3169 BP.  
XX  
AC AAS87641;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #23445.  
XX



KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR P-PSDB; ABG23454.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

PT

XX Claim 1; SEQ ID NO 23445; 103pp; English.

PS

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.27e-131	Length:	3169
Score:	1313.00	Matches:	293
Percent Similarity:	63.45%	Conservative:	23
Best Local Similarity:	58.84%	Mismatches:	57
Query Match:	38.58%	Indels:	125
DB:	5	Gaps:	9

US-09-932-678-2 (1-651) x AAS87641 (1-3169)

QY 1 MetAlaAProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20  
|||||

Db 1 ATGGCGCACCGCTGCTTCACACGCGTTTGCGGGGAGATGCGCGCGCTTCGCGCTCTGCA 60  
|||||

QY 21 ValLysLysLeuGlyAlaSerArgThrGlyLeSerAsnMetArgAlaLeuGluAsnAsp 40  
|||||

Db 61 GTCAAGATGCTGGGCGCGTCGAGGACTGGGATTCAAAATATGCGTGCATTAGAGAAATGAT 120  
|||||

QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
|||||

Db 121 TTTTTCAAATCTCCCCCAAGAAAAAAGTTCAGTTTGGTGGAAGTGTGACAGAAGTCTTG 180  
|||||

QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
|||||

Db 181 CTGAAGTACAAAAACGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTAGAT 240  
|||||

QY 81 ProAspIleLys-----AspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSer 98  
|||||

Db 241 CCAGACATAAAGGTGGTGTGCCCGCCATGCCGATGATCCTGAGGGAGGGAGGCCATT 300  
|||||

QY 99 Ile-----MetTyrLeuThrLysAspPheGluGln 108  
::: |||

Db 301 GTGGGCGAGTGTCCCTTTCATCCCTTCGFGGCTTTCAGTTGTCTCAAAACCAGAGTGGC 360  
|||||

QY 109 LeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGlu 128  
|||||

Db 361 CTCAGCCCGAAGGTTGCTAAGCTGCCTGCGTGGTTCGG----- 399  
::: |||

QY 129 TyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeu 148  
|||||

Db 400 -----CCCGTGGCCTGCCGCGCCACATAC 423  
|||||

QY 149 SerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAsp 168  
|||||

Db 424 AGCCCCAGCCTCTGCCAC----- 441  
|||||

QY 169 ValSerAspSerAspAspGluAspAspAsnLeuProAla-AsnPheAspThrCysHisAr 188  
::: |||

Db 442 -----AACCCCAACGCCCGCAGCTTCTGGCTCACTCAT-- 478  
|||||

QY 188 gAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLe 208  
::: |||

Db 479 -----CTGCTTTCAGG-----ACACCGTGGTTTCTCATGCCAATACT 516  
|||||

QY 208 uValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAs 228  
|||||

Db 517 GGTGGAATAATTCCATTGTTTCGAAATCAGAGAGAACACTGGAATGTACGTTCAATA 576  
|||||

QY 228 nLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIl 248  
|||||

Db 577 CTTACTAAGGATTAGTGATATATTTCCAACTTGAGGCATGAAATCTGGAGCTTATTAT 636  
|||||

QY 248 eGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGl 268  
|||||

Db 637 TGAAAAACTACTCAAGTTGGATGTGAATGCATCCCGGCGGAGTATTGAAGATGCTGAAGA 696  
|||||

QY 268 uThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGl 288  
|||||

Db 697 AACAGCAATCAAACCTTGTTGGTGGACAGATCCACGGAGGATTGTGTAATATGGTTAG 756  
|||||

QY 288 uAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHi 308  
|||||

Db 757 C----- 757  
|||||

QY 308 sProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspVa 328  
::: |||

Db 758 -----AGTTTATTATGAAGTGGAGATGAAGTTATC----- 790  
|||||

QY 328 lCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIl 348  
::: |||

Db 791 -----ATAAT 795  
|||||

QY 348 eAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMe 368  
::: |||

Db 796 CAAAGGTGGAAACAGCTAGTGCTG----- 820  
|||||

QY 368 tPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLy 388  
|||||

Db 821 -----CTCATCTTTGGATTGCGAGAGCATTTTGGAAACCTCTTTGGAA 864  
|||||

QY 388 sLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGl 408  
|||||

Db 865 AAAATTGCAGGATCCAAGTAATCCTGCCATCATCAGGAGGCTGCTGGAATATATATTGG 924  
|||||

QY 408 ySerPheLeuAlaArqAlaLysPheIleProLeuIleThrVallysSerCysLeuAspLe 428  
Db 925 AAGCTTTTGGCAAGAGCTAAATTTATTCCTTATTAAGTAAATCATGCCTAGATCT 984  
QY 428 uLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCy 448  
Db 985 TTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTGGGAACAAAGGCATTCTG 1044  
QY 448 sAspValAlaLeuHisGlyPropheTyrSerAlaCysGlnAlaValPheTyrThrPheVa 468  
Db 1045 CGATGTTGCTCTCCATGGACCATTTTACTAGCCTGCCAAGCTGTGTTCTACACCTTTGT 1104  
QY 468 lPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr 485  
Db 1105 TTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAGAAGGCGTCTGGTTC 1156

RESULT 9  
ABK47950  
ID ABK47950 standard; cDNA; 1989 BP.  
XX  
AC ABK47950;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Human transcriptional control factor ZFM1 isomer 25.63 cDNA.  
XX  
KW Human; transcriptional control factor ZFM1 isomer 25.63; gene; ss; HIV;  
KW malignant tumour; haemopathy; human immunodeficiency virus; cancer;  
KW immunological disease; inflammation; cytostatic; haemostatic; virucide;  
KW immunomodulatory; antiinflammatory; gene therapy.  
XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT CDS 1178..1879  
FT /\*tag= a  
FT /product= "Human transcriptional control factor ZFM1  
FT isomer 25.63"

XX WO200220588-A1.  
XX  
PD 14-MAR-2002.  
XX  
PF 02-JUL-2001; 2001WO-CN001127.  
XX  
PR 07-JUL-2000; 2000CN-00117050.  
XX (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.  
XX Mao Y, Xie Y;  
XX WPI; 2002-339796/37.  
XX P-PSDB; AAU77171.  
XX  
PT Human transcriptional control factor ZFM1 isomer 25.63 and encoding  
PT polynucleotide, used in diagnosis and treatment of malignant tumors,  
PT hemopathy, human immunodeficiency virus infection, immunological diseases  
PT and inflammation.

PS Claim 6; Page 30-31; 38pp; Chinese.  
XX  
CC The invention relates to the human transcriptional control factor ZFM1  
CC isomer 25.63 and the polynucleotide encoding it. The sequences of the  
CC invention are used in diagnosis and treatment of malignant tumours,  
CC haemopathy, human immunodeficiency virus (HIV) infection, immunological  
CC diseases and various inflammations. This sequence represents cDNA  
CC encoding the human transcriptional control factor ZFM1 isomer 25.63  
XX  
SQ Sequence 1989 BP; 559 A; 356 C; 430 G; 644 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1e-127 Length: 1989  
Score: 1280.50 Matches: 305

Percent Similarity: 40.90% Conservative: 5  
Best Local Similarity: 40.24% Mismatches: 21  
Query Match: 37.63% Indels: 428  
DB: 6 Gaps: 5  
US-09-932-678-2 (1-651) x ABK47950 (1-1989)  
QY 126 ValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145  
Db 1 GTGGAAGAGTATTGGCTTTTCTTGGTAATCTTGTAATCAGCATAGACTGTTTTCCTCAGA 60  
QY 146 ProCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGly 165  
Db 61 CCGTGTCTCAGCATGATGCTTCCCATTTTGTGCCTCCCGAGTGATCATTAAGGAAGGC 120  
QY 166 AspValAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThr 185  
Db 121 GATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCTCTGCAAAATTTTGACACA 180  
QY 186 CysHisArgAlaLeuGlnIleIleAlaArgTyrValProSer----- 199  
Db 181 TATCACAGAGCCTTGCAAAATAATAGCAAGGTATGTACCATCATGAGTATATACTTTTCCTTA 240  
QY 199 ----- 199  
Db 241 TTTTGAATGTTTAAATTTCTCAAGAAAATGTAAATCAATTAGTAAAAATATAAAATGTAA 300  
QY 199 ----- 199  
Db 301 TAGTATAAAGCTTGAGTCTTACATTGCATGTTTTTTTTTTTGTATCCACTTGAGGAAAC 360  
QY 199 ----- 199  
Db 361 ATTACATTCTACAAAAGTGGCATTTCCATTTTCTATTATTCTCTTTAACTGTTTTTCA 420  
QY 199 ----- 199  
Db 421 AAGTTCGTATGCAGATTCTCCCCCAATTTTGTATGGTGGTTGCGGTTTCTTTTATCTT 480  
QY 199 ----- 199  
Db 481 CAACAGATATGCTATCCAAAATTTTTCAGTGAGAAACCCCTGGGTGTGTTGTGTCATGC 540  
QY 199 ----- 199  
Db 541 CATATGAATAAAATTCACATTCTAAGAAAAAGCTTTTCAGGTTTGTGCGTTTCTTTTGA 600  
QY 199 ----- 199  
Db 601 GGGGTGGATTCTAGTTCCTCTGTCTGTGATTATTGTGTTAACTTAAAAAATCCAAC 660  
QY 199 ----- 199  
Db 661 TGATTATTTTCTTCTTTAAATAATAATATACATGTGTAGTGGGAAATGTCAGCAAAAG 720  
QY 199 ----- 199  
Db 721 TGCTGTATGTTTCTCTGGGAGAGAGCTCCCTCTTTGATTTGCTGTGTATATCAGAGTT 780  
QY 199 ----- 199  
Db 781 AACAGAAGCTTATTTTCTTAAGTTGTATAGACTTTTCTCAGAAGCTATACATTGTAAGT 840  
QY 199 ----- 199  
Db 841 TCCAGTTCTGGCGGCGGGTGGCTCACACCTGTAAATCCAGCACTTTGGGAGGCTGAG 900  
QY 199 ----- 199  
Db 901 GCGGGCGGATCACCTGAGGTCGGGAGTTCGAGACTAGCCTGACCAACATGGAGAAACCCC 960  
QY 199 ----- 199

Db 961 GTCTCTACTAAAAATACAAAATTAGGTGGCGTGGTGGCGCATGCTGCTAATCCAGCTG 1020  
QY 199 ----- 199  
Db 1021 TTTGGGGGCTGAGCGAGGAGAAATCGCTTGAACCCGGGAGGCAGAGGTTACAGTGAGCCG 1080  
QY 199 ----- 199  
Db 1081 AGATTGTGCCACTGCACTCCAGCCTGGGCAACAAGAGTGAAACTCCGTCTCAAAAAACAA 1140  
QY 200 -----ThrProTrpPheLeuMetProIleLeuValGluLysPh 212  
Db 1141 AAAAAAGTTCAGTTCTTTGAGGGCACCGTGGTTTCTCATGCCAATFACTGGTGGAAAAATT 1200  
QY 212 eProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIl 232  
Db 1201 TCCATTTGTTTCGAAAATCAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGAT 1260  
QY 232 eSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLe 252  
Db 1261 TAGTGATATATTTTCCAACCTTGAGGCATGAAATCTCGAGCTTATTATTGAAAACTACT 1320  
QY 252 uLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGl 272  
Db 1321 CAAGCTGGATGTAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAATCA 1380  
QY 272 nThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluTh 292  
Db 1381 AACTTGTGGTGACAGATTCCACGGAAGGATTGTTAATATG----- 1423  
QY 292 rGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGl 312  
Db 1423 ----- 1423  
QY 312 uArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAs 332  
Db 1423 ----- 1423  
QY 332 pGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAs 352  
Db 1423 ----- 1423  
QY 352 pLysLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCy 372  
Db 1423 ----- 1423  
QY 372 sSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAs 392  
Db 1424 -----GGATTCCGAGAGGCATTTTGGAAACATCTTTGGAAAAACTTGCAGGA 1470  
QY 392 pProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAl 412  
Db 1471 TCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATATATATTGGAAGCTTTTGGC 1530  
QY 412 aArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTr 432  
Db 1531 AAGAGCTAAATTTATTCTCTTATTACTGAAAAACCATGCTAGATCTTTTGGTTAACTG 1590  
QY 432 pLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLe 452  
Db 1591 GCTGCACATATACCTTAATAACCAAGGATTCGGGAACAAAGGCATCTGCGATGTTGCTCT 1650  
QY 452 uHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLy 472  
Db 1651 CCATGGACCATTTTACTCAGCTGCCAAGTGTGTTCTACACCTTTGTTTTAGACACAA 1710  
QY 472 sGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGl 492  
Db 1711 GCAGCTTTTGAGCGGAAACCTGAAAGAGGTTGCAGTATCCTCAGAGTCTGAATTTTGA 1770  
QY 492 uArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPh 512  
Db 1771 GCGGATAGTGATGAGCCAGCTAAATCCCCGTAAGATTTCCTGCCCTCAGTGGTTAACTT 1830

QY 512 ePheAlaAlaIleThrAsnLysTyr-GlnLeuValPheCysTyrThrIleIleGluArgA 532  
Db 1831 TTTTGCTGCAATCACA---AAGATGAAGACTTGTG----- 1862  
QY 532 snAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleC 552  
Db 1863 -----GATATGGATGGTGGTGATGGT-----T 1884  
QY 552 ysThrAsnProLeuAspThrPhePheProPheAspProCysValleuLys 568  
Db 1885 GCACAACAAT-ATCAATTTATTTATACCACCTGAACCGTGCACTTCAAAA 1933  
RESULT 10  
AAS92254  
ID AAS92254 standard; cDNA; 3175 BP.  
XX AAS92254;  
AC AAS92254;  
XX 13-FEB-2002 (first entry)  
DT DNA encoding novel human diagnostic protein #28058.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
KW Homo sapiens.  
OS WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US008631.  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI MPI; 2001-639362/73.  
DR P-PSDB; ABG28067.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX Claim 1; SEQ ID NO 28058; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 3175 BP; 811 A; 849 C; 779 G; 736 T; 0 U; 0 Other;

Alignment Scores:

Pred. NO.: 3.01e-119 Length: 3175  
Score: 1205.00 Matches: 288  
Percent Similarity: 61.90% Conservative: 24  
Best Local Similarity: 57.14% Mismatches: 61  
Query Match: 35.41% Indels: 131  
DB: 5 Gaps: 9

US-09-932-678-2 (1-651) x AAS922254 (1-3175)

Qy	1	MetAlaLaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAspSerSerAla	20
Db	1	ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGAGATGCGCGCTTCGCTCTGCA	60
Qy	21	VallLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp	40
Db	61	GTCAAGATGCTGGCGCGTCGAGGACTGGGATTTCAAATATGCTGTCATTAGAGAATGAT	120
Qy	41	PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu	60
Db	121	TTTTTCAATTCTCCCCAGAAAAAACTGTCAGTTTGTGTGAACTGTGACAGAGTCITG	180
Qy	61	LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp	80
Db	181	CTGAAGTACAAAAACGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT	240
Qy	81	ProAspIleLys-----AspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSer	98
Db	241	CCAGACATAAAGGTGTGTGCCCCCGCATGCCGATGATCCTGAGGGAGGGAGGCCATT	300
Qy	99	Ile-----MetTyrLeuThrLysAspPheGluGln	108
Db	301	GTGGCGAGTGTGCCTTTTCATCCCTTCGTGGCTTTTCAGTTGTGCTCAAACACAGATGGC	360
Qy	109	LeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGlu	128
Db	361	CTCAGCCCCGAAGTTGTCTAAGCTGCCTGCGTGGTTCGG-----	399
Qy	129	TyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeu	148
Db	400	-----CCCGTGGCTGCGCGGCCACATAC	423
Qy	149	SerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAsp	168
Db	424	AGCCCCAGCCTCTGCCAC-----	441
Qy	169	ValSerAspSerAspGluAspAspAsnLeuProAla-AsnPheAspThrCysHisAr	188
Db	442	-----AACCCCCAACGCCAGCGCAGCTTCTGGCTCACTCAT--	478
Qy	188	gAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLe	208
Db	479	-----CTGCTTTGCAGG-----ACACCGTGGTTTCTCATGCCAATACT	516
Qy	208	uValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAs	228
Db	517	GGTGAAAAAATTCCATTTGTCGAAAAATCAGAGAAACACTGGAATGTTACGTTCATAA	576
Qy	228	nLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIl	248
Db	577	CTTACTAAGGATTAGTGATATTTTCCAACCTTGAGGCATGAAATTCCTGAGCTTATTAT	636
Qy	248	eGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGl	268
Db	637	TGAAAAAACTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGA	696
Qy	268	uThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGl	288
Db	697	AACAGCAAAATCAAACTTGTGTGGGACAGATTCCACGGAAGGATGTGTAATATGGTTAG	756
Qy	288	uAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHi	308

Db	757	C-----	757
Qy	308	sProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspVa	328
Db	758	-----AGTTTATTAAATGAAAGTGGAGATGAAGTTTATC-----	790
Qy	328	lCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIl	348
Db	791	-----ATAAT	795
Qy	348	eAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMe	368
Db	796	CAAAGGCTGGAAACAGCTAGTGCTG-----	820
Qy	368	tPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLy	388
Db	821	-----CTCATCTTTGGATTGCGAGGCAATTTTGGAAACCTCTTTGGAA	864
Qy	388	sLysLeuGln-AspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleG	408
Db	865	AAAACTGCAGGGATCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGACATTATTTG	924
Qy	408	lySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspL	428
Db	925	CAAGCTTTTGGCAAGAGCTACATTTATTCTCTTATTACTGTAAAACCATGCCTAGATC	984
Qy	428	euLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSer-GlyThrLysAlaPhe	447
Db	985	TTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGGAACAAAGGCATT	1044
Qy	448	CysAspValAla-LeuHisGly-ProPheTyrSerAlaCys-GlnAlaValPheTyrThr	466
Db	1045	TGCGATGTGCCTCTCCATGGACCCCAATTTTACTCAGCCTGCCCAAGCTGTGTCTACACC	1104
Qy	467	PheValPheArg-HisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr	485
Db	1105	TTTGTTTTAGAACACACAGCAGCTTTTGACGGGAAACCTGAAAGAAGGCGTGTGGTTC	1162

RESULT 11

AAS87636  
ID AAS87636 standard; cDNA; 1017 BP.  
XX  
AC AAS87636;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #23440.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG23449.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX PS Claim 1; SEQ ID NO 23440; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1017 BP; 301 A; 214 C; 230 G; 272 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.06e-117 Length: 1017  
Score: 1181.00 Matches: 249  
Percent Similarity: 55.31% Conservative: 1  
Best Local Similarity: 55.09% Mismatches: 4  
Query Match: 34.70% Indels: 198  
DB: 5 Gaps: 2

US-09-932-678-2 (1-651) x AAS87636 (1-1017)

QY	30	GlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsnSerProProArgLysThr	49
Db	7	GGGATTTCAAATATGCGTGCATTAGAGATGATTTTTCATTCTCCCCAAGAAACT	66
QY	50	ValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsn	69
Db	67	GTTCGGTTTGGTGGAACTGTGACAGAACTGTGCTGAAGTACAAAAGGGTGAAACAAAT	126
QY	70	AspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIle	89
Db	127	GACTTTGAGTTGTTGAAGAACCAAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATC	186
QY	90	AsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeu	109
Db	187	AACTGGCTGTAGAAATCCGTTCTTCTGTCTATGTAAGTAAAGACCTTGAGCAACTT	246
QY	110	IleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyr	129
Db	247	ATCAGTATTATA-----	258
QY	130	LeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSer	149
Db	258	-----	258
QY	150	MetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAspVal	169
Db	258	-----	258
QY	170	SerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAla	189
Db	258	-----	258
QY	190	LeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeuVal	209
Db	258	-----	258

QY	210	GluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeu	229
Db	259	-----TTGGAAATGTTACGTTTCATAACTTA	282
QY	230	LeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleGlu	249
Db	283	CTAAGGATTAGTGATATTTTCCAACCTTGAGGCATGAAATTTGGAGCTTATTATTGAA	342
QY	250	LysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThr	269
Db	343	AAACTACTCAAGCTGGATGTAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACA	402
QY	270	AlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAsp	289
Db	403	GCAAATCAAACTTGTGTGGACAGATTCCACGGAAGGATTGTTTAATATG-----	453
QY	290	GluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisPro	309
Db	453	-----	453
QY	310	ValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCys	329
Db	453	-----	453
QY	330	TyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsn	349
Db	453	-----	453
QY	350	IlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPhe	369
Db	453	-----	453
QY	370	TyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLys	389
Db	454	-----GGATTGCGACAGGCGATTTTGGAAACATCTTTGGAAAAAC	492
QY	390	LeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySer	409
Db	493	TTGCAGGATCCAAGTAATCTGCCATCATCAGGCAGGCTGCTGGAAATATATTTGGAAGC	552
QY	410	PheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeu	429
Db	553	TTTTTGGCAAGAGCTAAATTTATTTCTCTTATTACTGTAAACCATGCTTAGATCTTTTG	612
QY	430	ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAsp	449
Db	613	GTTAACCTGGCTGCACATATACCTTAATAACCAGGATTTCGGGAACAAAGGCATTCTGCGAT	672
QY	450	ValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPhe	469
Db	673	GTTGCTCTCCATGGACCATTTTACTCAGCTGCCAAGCTGTGTCTACACCTTTGTTTTT	732
QY	470	ArgHisLysGlnLeuLeuSerGlyAsnLeuLysGlu	481
Db	733	AGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAA	768

RESULT 12

ABA83038

ID ABA83038 standard; DNA; 1461 BP.

XX AC ABA83038;

XX AC ABA83038;

DT 05-FEB-2002 (first entry)

XX Human transcription factor TRFX-65 coding sequence.

DE Human; transcription factor; TRFX; cell proliferative disease;

XX Human; transcription factor; TRFX; cell proliferative disease;

KW autoimmune disease; inflammation; neurological disease;

KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;

KW neuroprotective; antiinflammatory; gene therapy; ds.

XX Homo sapiens.

OS

XX WO200172777-A2.  
XX  
PD 04-OCT-2001.  
XX  
PF 13-MAR-2001; 2001WO-US008117.  
XX  
PR 13-MAR-2000; 2000US-0188986P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
PI Hillman JL, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;  
PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;  
PI Reddy R;  
XX  
DR WPI; 2001-570896/64.  
DR P-PSDB; ABB50214.  
XX  
XX Novel transcription factor polypeptides, used to treat diseases  
PT associated with altered activity and expression of TRFX, and to screen  
PT for agents capable of modulating its activity.  
PS Claim 11; Page 299; 327pp; English.  
XX  
CC The present sequence is the coding sequence for a human transcription  
CC factor. The transcription factor and its coding sequence are useful in  
CC the diagnosis, treatment and prevention of diseases associated with  
CC altered expression of the transcription factor e.g. cell proliferative,  
CC autoimmune/inflammatory, neurological and developmental disorders. A  
CC number of specific disorders/diseases are given in the specification,  
CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
CC disease, stroke, and viral, bacterial, fungal and protozoal infections  
XX  
SQ Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 5.24e-111 Length: 1461  
Score: 1124.00 Matches: 256  
Percent Similarity: 55.44% Conservative: 14  
Best Local Similarity: 52.57% Mismatches: 52  
Query Match: 33.03% Indels: 166  
DB: 5 Gaps: 10  
  
US-09-932-678-2 (1-651) x ABA83038 (1-1461)  
  
QY 195 ArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPhe 214  
Db 263 AAATACAAAATTAGGGCACCGTGGTTCTCATGCCAATACTAGTGGTGGAAAAATTTCCATT 322  
  
QY 215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerVal 234  
Db 323 GTTCGAAAATCAGAGAACACTTGGAAATGTACGTTCAATAACTTACTAAGGATTAGTGTA 382  
  
QY 235 TyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLysLeu 254  
Db 383 TATTTTCCAACCTTGAGGCATGAAATTTCTGGAGCTTATTATTGAAAACTACTCAAGCTG 442  
  
QY 255 AspValAsnAlaSerArgGlnGlyIleGluAspAlaGluThrAlaThrGlnThrCys 274  
Db 443 GATGTGAATGCATCCCGCAGGGGTATTGAAGATGCTGAAGAAACAGCAATCAAACTTGT 502  
  
QY 275 GlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluThrGluHis 294  
Db 503 GGTGGACAGATTCCACGGAAGGATTGTTTAATATG----- 538  
  
QY 295 GluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeu 314  
Db 538 ----- 538

QY 315 AspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLys 334  
Db 538 ----- 538  
  
QY 335 ValAspAsnGlyLysThrLysAspLeuTyrArgAspIleuIleAsnIlePheAspLysLeu 354  
Db 538 ----- 538  
  
QY 355 LeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPhe 374  
Db 538 ----- 538  
  
QY 375 LysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSer 394  
Db 539 -----GGATTGCGAGAGGCATTTTGGAACATCTTTGGAAAAACCTTGCAAGATCCAAGT 592  
  
QY 395 AsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAla 414  
Db 593 AATCCTGCCATCATCAGGCAGGCTGCTGGAATTTATATTGGAAGCTTTTGGCAAGAGCT 652  
  
QY 415 LysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHis 434  
Db 653 AAATTATTCTCTTATTACTGTAAAAACCATGCCTAGATCTTTTGGTTAACTGGCTGCAC 712  
  
QY 435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454  
Db 713 ATATACCTTAATAACAGGATTCGGGAACAAAGGCATCTCGATGTGTCTCTCCATGGA 772  
  
QY 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474  
Db 773 CCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTGTGTTTAGACACAAGCAGCTT 832  
  
QY 475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494  
Db 833 TTGAGCGGAAACCTGAAAGAGGTTTGCAGTATCCTCAGAGTCTGAATTTTGAGCGGATA 892  
  
QY 495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla 514  
Db 893 GTGATGAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTGTGCT 952  
  
QY 515 AlaIleThrAsnLysTyr-GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnAr 534  
Db 953 GCAATCACA---AAGATGAAGACTTTGTG----- 977  
  
QY 534 gGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAs 554  
Db 978 -----GATATGGATGGTGGTGATGGT-----TGCACAAC 1006  
  
QY 554 nProLeuAspThrPhePheProPheAspProCysValLeuLys----- 568  
Db 1007 AAT-ATCAATTTATTTTATACCACCTGACCTGACCTTCAAAATGGTTAAGATGGTAAGT 1065  
  
QY 569 -----ArgSerLysLysPheIleAspPr 576  
Db 1066 TTTATGTGTGTATTTTACCATAATAAAAAAAATTTGAGAGGGAAAAACAGTCTGCCTCC 1125  
  
QY 576 oIleTyrGlnValTrpGlu-----AspMetSerAlaGluGl 588  
Db 1126 ACTTTTG---ATATGGGACTGCTAACATCTTCCACCTTCCGCTCTCCCTTCTGCCCCACA 1182  
  
QY 588 uLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAspAspPh 608  
Db 1183 TCTGGCAAGCTAAGAAAGCCT----- 1204  
  
QY 608 eLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer----- 625  
Db 1205 -----GCTGCTCTCTCCTCTGGCACCAGCTGGGAATTCATACCCCAACAGCCCTA 1254  
  
QY 626 -----SerPheAspThrHisPheArgSerProSerSerValGlySerProProValle 644  
Db 1255 GCCTTCCACAGACCCACATTTTCATCCCATCCCATCCCATCCCATCCC----- 1306



QY 644 uTyrMetGlnProSerPro 650  
Db 1307 -----CATCCCATCCC 1318

RESULT 13  
AAS92255  
ID AAS92255 standard; cDNA; 2410 BP.

XX AAS92255;  
AC AAS92255;  
XX 13-FEB-2002 (first entry)  
DT  
XX DNA encoding novel human diagnostic protein #28059.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
KW Homo sapiens.  
OS  
XX WO200175067-A2.  
PN 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US008631.  
PF  
XX 31-MAR-2000; 2000US-00540217.  
PR  
XX 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG28068.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX Claim 1; SEQ ID NO 28059; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2.89e-105 Length: 2410  
Score: 1074.00 Matches: 220  
Percent Similarity: 95.26% Conservative: 1  
Best Local Similarity: 94.83% Mismatches: 4  
Query Match: 31.56% Indels: 7  
DB: 5 Gaps: 0

US-09-932-678-2 (1-651) x AAS92255 (1-2410)

QY 427 AspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAla 446  
Db 1 GATCTTTTGGTTAACTGGCTGCATATACCTTAATAACCAAGGATTCGGGAACAAGGCA 60

QY 447 PheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThr 466  
Db 61 TTCTGCGATGTGCTCTCCATGAGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACC 120

QY 467 PheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeu 486  
Db 121 TTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTT 180

QY 487 GlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCys-Le 506  
Db 181 CAGAGTCTGAATTTTGAGTGGATAGTGATGAGCCAGCTAAATCCCTGAAGATATGGCTT 240

QY 506 uProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTy 526  
Db 241 GCCCTCAGTGGTTAACTTTTGTGCTGCAATCAAAATAAGTACCAGCTCGTCTTCTGCTA 300

QY 526 rThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGl 546  
Db 301 CACCATCATCGAGAGGAACAATCGCCAGATGTCGCCAGTCATTAGGAGTACCCTGGAGG 360

QY 546 YAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPhe-AspProCysV 566  
Db 361 AGACTCAGTGCAGACCTGCACAAACCCACTGGACACCTTCTTCCCTTTGGATCCCTGTG 420

QY 566 alLeuLysArgSerLysLysPhe-IleAspProIle-Tyr-GlnValTrp-GluAspMet 584  
Db 421 TGCTGAAGAGGTCAAAGAAATTCATTGATCTATTTTATCCAGGTATGGGAAGACATG 480

QY 585 SerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGlu 604  
Db 481 AGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAGGACATAGTGAAGATGAA 540

QY 605 AspAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrPro 624  
Db 541 GATGATGACTTCTGAAAGCGAAGTGGCCCCAGAAATGATACCGTGATTTGGGATCACCCA 600

QY 625 SerSerPheAspThrHisPheArgSerProSerSerSerValGlySerProProValLeu 644  
Db 601 AGCTCCTTTGACACGCAATTTCCGAAGTCTCTCAAGTAGTGTGGGCTCCCCACCCGTGTG 660

QY 645 Tyr-MetGlnProSerProLeu 651  
Db 661 TACTGCAACCCAGTCCCTC 682

RESULT 14  
ABS64829  
ID ABS64829 standard; DNA; 2493 BP.  
XX  
AC ABS64829;  
XX  
DT 15-NOV-2002 (first entry)  
XX  
DE Human tumour suppressor gene #3.  
XX  
KW Human; cancer; gene; ds; tumour suppressor gene; breast cancer; lymphoma;  
KW adenocarcinoma; leukaemia; melanoma; sarcoma; teratocarcinoma;  
KW prostate; cervix; liver; ovary; adrenal gland; heart; brain; lung; colon;  
KW placenta; skeletal muscle; synovial membrane; tonsil; kidney; uterus;  
KW skin; cytostatic.

XX  
OS Homo sapiens.  
XX  
PN WO200264775-A1.  
XX  
PD 22-AUG-2002.  
XX

PF 12-FEB-2002; 2002WO-AU000137.  
XX  
PR 12-FEB-2001; 2001AU-00003052.  
PR 12-FEB-2001; 2001AU-00003053.  
PR 12-FEB-2001; 2001AU-00003054.  
XX  
PA (BION-) BIONOMICS LTD.  
XX  
PI Callen DF, Powell JA, Krennmidiotis G, Gardner AE, Whitmore SA;  
XX  
DR WPI; 2002-657597/70.  
DR P-PSDB; ABG79517.  
XX  
PT New tumor suppressor genes identified at 16q24.3, useful for identifying  
PT or obtaining full-length human genes involved in the tumorigenic process,  
PT or in diagnosing or treating cancer (e.g. breast or cervix cancer)  
PT through gene therapy.  
XX  
PS Claim 10; Page 76-79; 218pp; English.  
PS  
XX  
CC The invention relates to a tumor suppressor gene and the protein it  
CC encodes. The genes are useful for identifying and/or obtaining full-  
CC length human genes involved in the tumorigenic process. The genes the  
CC polypeptides and antibodies to the polypeptides are useful in diagnosing  
CC cancer, in establishing the prognosis of a patient diagnosed with cancer,  
CC or in treating cancer through gene therapy. In particular, the cancer is  
CC breast cancer. The drugs that restore tumor suppressor activity are  
CC useful for treating breast cancer or for manufacturing a medicament for  
CC the treatment of breast cancer. The cancer may also be adenocarcinoma,  
CC leukaemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, or  
CC cancer of the prostate, cervix, liver, ovary, adrenal gland, heart,  
CC brain, lung, placenta, skeletal muscle, synovial membrane, tonsil,  
CC kidney, colon, uterus or skin. This sequence represents a human tumour  
CC suppressor gene of the invention  
XX  
SQ Sequence 2493 BP; 554 A; 700 C; 761 G; 478 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 3.06e-97 Length: 2493  
Score: 1000.00 Matches: 188  
Percent Similarity: 97.45% Conservative: 3  
Best Local Similarity: 95.92% Mismatches: 5  
Query Match: 29.39% Indels: 0  
DB: 6 Gaps: 0

US-09-932-678-2 (1-651) x ABS64829 (1-2493)

QY 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393  
Db : : : : :  
703 TATCGCAGGGATTGCGAGAGGCATTTTGGAAACATCTTTGGAAAAATTCAGGATCCA 762

QY 394 SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413  
Db : : : : :  
763 AGTAATCCTGCCATCATCAGCAGGCTGCTGGAAATTATATTGGAAGCTTTTGGCAAGA 822

QY 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu 433  
Db : : : : :  
823 GCTAAATTTATTCTCTTTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTG 882

QY 434 HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453  
Db : : : : :  
883 CACATATACCTTAATAACAGGATTCGGGAACAAAGGCATCTGCAATGTGCTCTCCAT 942

QY 454 GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473  
Db : : : : :  
943 GGACCATTTTACTCAGCCTGCCAAGCTGTGTCTACACCTTTGTTTATAGACACAAGCAG 1002

QY 474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg 493  
Db : : : : :  
1003 CTTTGTAGCGGAAACCTGAAACAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGACGG 1062

QY 494 IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe 513  
Db : : : : :  
513

Db 1063 ATAGTGATGAGCCAGCTAAATCCCTGAAGATTTCCTGCCCTCAGTGGTTAACTTTT 1122

QY 514 AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533  
Db : : : : :  
1123 GCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGTACACCATCATCGAGGAACAAT 1182

QY 534 ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr 553  
Db : : : : :  
1183 CGCCAGATGCTGCCAGTCATTAGGAGTACCACCTGGAGGAGACTCAGTCGAGACCTGCACA 1242

QY 554 AsnProLeuAspThrPhePhePropheAspProCysValLeuLysArg 569  
Db : : : : :  
1243 AACCCACTGGACACCTTCGTCCTTTCATCCCTGTGTGCTGAAGAGG 1290

RESULT 15  
ADA52976  
ID ADA52976 standard; cDNA; 2493 BP.  
XX  
AC ADA52976;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human coding sequence, SEQ ID 544.  
XX  
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FN EP1293569-A2.  
XX  
PD 19-MAR-2003.  
XX  
PF 21-MAR-2002; 2002EP-00006586.  
XX  
PR 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
XX  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX  
DR WPI; 2003-395539/38.  
DR P-PSDB; ADA54615.  
XX  
PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 1; SEQ ID NO 544; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 2493 BP; 554 A; 700 C; 761 G; 478 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 3.06e-97 Length: 2493  
Score: 1000.00 Matches: 188  
Percent Similarity: 97.45% Conservative: 3  
Best Local Similarity: 95.92% Mismatches: 5  
Query Match: 29.39% Indels: 0  
DB: 7 Gaps: 0

US-09-932-678-2 (1-651) x ADA52976 (1-2493)

QY	374	PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTirLysLysLeuGlnAspPro	393
Db	703	TTTTTTTATCGCAGGGATTGCGAGAGGCAATTTTGGAAACATCTTTGGAAAAAATTCAGGATCCA	762
QY	394	SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg	413
Db	763	AGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATTATATTGGAAGCTTTTGGCAAGA	822
QY	414	AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTirPLeu	433
Db	823	GCTAAATTTATTCCTCTTATTACTGTAAACCAATGCTAGATCTTTGGTTAACTGGCTG	882
QY	434	HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis	453
Db	883	CACATATACCTTAATAAACCAAGGATTTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCAT	942
QY	454	GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln	473
Db	943	GGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAGCAG	1002
QY	474	LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg	493
Db	1003	CTTTTGAGCGGAAACCTGAAAGAGGTTTGCAAGTATCTTCAGAGTCTGAATTTTGAGCGG	1062
QY	494	IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe	513
Db	1063	ATAGTGATGAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTT	1122
QY	514	AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn	533
Db	1123	GCTGCAATCACAAATAAGTACCAGCTCGTCTCTCTGCTACACCATCATCGAGAGGAACAAT	1182
QY	534	ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr	553
Db	1183	CGCCAGATGCTGCCAGTCAATTAGGAGTACCAGTGGAGGAGACTCAGTGCAGACCTGCACA	1242
QY	554	AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArg	569
Db	1243	AACCCACTGGACACCTTCGTCCCTTTTGATCCCTGTGTGTGAAGAGG	1290



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 06:25:37 ; Search time 6976 Seconds  
(without alignments)  
4044.770 Million cell updates/sec

Title: US-09-932-678-2  
Perfect score: 3403  
Sequence: 1 MAAPLLHLRLPGDAAASSSA.....RSPSSSVGSPVLYMQPSPL 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09932678/runat\_15072004\_103603\_5206/app\_query.fasta\_1.839  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rgc -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09932678 @CGN 1 1 4545 @runat\_15072004\_103603\_5206 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3403	100.0	2068	9	AF227156	AF227156 Homo sapi
2	3398	99.9	3756	9	BSA272050	AJ272050 Homo sapi
3	3391	99.6	2040	6	AX036047	AX036047 Sequence
4	3391	99.6	3709	9	BC036182	BC036182 Homo sapi
5	2372.5	69.7	3166	10	BC034110	BC034110 Mus muscu
6	2039	59.9	135873	2	AC012547	AC012547 Homo sapi
7	2039	59.9	192815	9	AC013444	AC013444 Homo sapi
8	1973	58.0	1423	6	AX427479	AX427479 Sequence
9	1124	33.0	1461	6	AX274907	AX274907 Sequence
10	1003	29.5	2627	9	BC006441	BC006441 Homo sapi
11	1000	29.4	2493	6	AX713860	AX713860 Sequence
12	1000	29.4	2493	9	AK055742	AK055742 Homo sapi
13	995.5	29.3	2373	3	AK114871	AK114871 Ciona int
14	991.5	29.1	2243	10	BC055781	BC055781 Mus muscu
15	986	29.0	3645	9	AK126166	AK126166 Homo sapi
16	924	27.2	1168	9	BC009198	BC009198 Homo sapi
17	783	23.0	2112	3	AY061123	AY061123 Drosophil
18	712	20.9	437	6	BD073832	BD073832 5'EST of
19	554.5	16.3	37621	8	SPAC18G6	Z68198 S.pombe chr
20	505	14.8	341	6	BD072397	BD072397 Secreted
21	499.5	14.7	1671	6	AX489470	AX489470 Sequence
22	476.5	14.0	2165	8	AY096642	AY096642 Arabidops
23	474	13.9	1873	8	AY142672	AY142672 Arabidops
24	456.5	13.4	1815	6	AX507509	AX507509 Sequence
25	450.5	13.2	2354	8	AK101144	AK101144 Oryza sat
26	424.5	12.5	2672	8	SCRN3GEN	Z71927 S.cerevisia
27	424.5	12.5	2843	8	SCYL125W	Z28125 S.cerevisia
28	424.5	12.5	75335	8	BX842627	BX842627 Neurospor
29	420.5	12.4	1884	6	AX596826	AX596826 Sequence
30	420.5	12.4	1884	6	AX821254	AX821254 Sequence
31	420.5	12.4	1884	6	AX832284	AX832284 Sequence
32	403	11.8	191598	2	AC144362	AC144362 Gorilla g
33	403	11.8	221030	2	AC145403	AC145403 Gorilla g
34	398	11.7	131839	9	AC112166	AC112166 Homo sapi
35	398	11.7	157631	2	AC142079	AC142079 Homo sapi
36	398	11.7	170613	9	AC106788	AC106788 Homo sapi
37	398	11.7	175691	2	AC007615	AC007615 Homo sapi
38	398	11.7	185887	2	AC133553	AC133553 Homo sapi
39	398	11.7	207997	9	AC009093	AC009093 Homo sapi
40	393.5	11.6	183586	9	AC092562	AC092562 Papio ham
41	392	11.5	129473	2	AC040158	AC040158 Homo sapi
42	392	11.5	143783	9	AC139256	AC139256 Homo sapi
43	392	11.5	184471	9	AC136026	AC136026 Homo sapi
44	392	11.5	184891	9	AC092137	AC092137 Homo sapi
45	392	11.5	189770	2	AC139281	AC139281 Homo sapi

ALIGNMENTS

AF227156	AF227156	2068 bp	mRNA	linear	PRI 29-APR-2000
LOCUS					
DEFINITION	Homo sapiens RRN3 mRNA, complete cds.				
ACCESSION	AF227156				
VERSION	AF227156.1 GI:7670099				
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2068)				
AUTHORS	Moorefield,B., Greene,E.A. and Reeder,R.H.				
TITLE	RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)				
MEDLINE	20243763				
PUBMED	10758157				
REFERENCE	2 (bases 1 to 2068)				
AUTHORS	Moorefield,B., Greene,E.A. and Reeder,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA				
FEATURES	Location/Qualifiers				
source	1..2068				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/chromosome="16"				
	/map="16p12"				
	69..2024				
	/note="similar to yeast RNA polymerase I transcription factor Rrn3"				
	/codon_start=1				
	/product="RRN3"				
	/protein_id="AAF66160.1"				
	/db_xref="GI:7670100"				
	/translation="MAAPLLHTRLPGDAASSAVKKGASRTGISMRALENDFNFSPPKTVRFQGTVEVLLKYGKETNDFFELKNQLLDPDIDKDDQIINWLLFRSSIMYLTKDQELISIIIRLPWLNRSTVVEEYLAFLGNLVSAQTVFLRPLCSMTASHFVPPRVIIKEGDDVSDSDDENDLPANFDTPCHRALQIARIYVPSPTWFLMPILVEKFPFVRKSERTLECYVHNLRLRISVYPTLRHEILELIEKLLKLDVNASRQIEDAETATQTCGGTDSTGLFNWDEDETEHETKAGPERLDQWHPVAERLDILMSLVLSYMKDVCYVDGKVDNGKTKDLYRDLINIFDKLLPTHASCHVQFMFYLCSPKLGFAEAFLEHLMKKLQDPSNPALIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNLHLYLNNDQSGTKAFCDVALHGPFYSACQAVFYTFVRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPSVWNFFAALTNKYQLVFCYTIERNNRQMLPVIIRSTAGDSVQICTNPLDFFFPDPCVLKRSKKFIDPIYQVWEDMSAELEQEFKPKMKDIVEDDDDFLKGEVPQNDTVIGITPSSFDITFRSPSSSVGPPVLYMQPSPL"				
ORIGIN					
Alignment Scores:					
Pred. No.:	2.05e-316	Length:	2068		
Score:	3403.00	Matches:	651		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	9	Gaps:	0		
US-09-932-678-2 (1-651) x AF227156 (1-2068)					
QY	1	MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerSerAla	20		
Db	69	ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATCGGCCGCTTCGTCTCTGCA	128		
QY	21	VallysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp	40		
Db	129	GTTRAGAAAGCTGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGATGAC	188		
QY	41	PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu	60		
Db	189	TTTTTTCATTCTCCCCAAGAAAACTGTTCCGTTTGGTGGAACTGTGACAGAACTCTTG	248		
QY	61	LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp	80		

Db	249	CTGAAGTACAAAAGGGTGAACAATAATGACTTTGAGTTGTTGAAGAACCACTGTTAGAT	308		
QY	81	ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet	100		
Db	309	CCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAAATTCGGTCTTCTATCATG	368		
QY	101	TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn	120		
Db	369	TACTTGACAAAAGACTTTGAGCAACTATCAGTATTATATAAGATTGCCTGGTTGAAT	428		
QY	121	ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln	140		
Db	429	AGAAGTCAAACAGTAGTAGGAAGAGTATTGGCTTTTCTGGTAATCTTGATCAGCACAG	488		
QY	141	ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal	160		
Db	489	ACTGTTTTCCTCAGACCGGTGCTCAGCATGATTGCTTCCCATTTTGCCTCCCCGAGTG	548		
QY	161	IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAsnLeuPro	180		
Db	549	ATCATTAAAGGAAGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCT	608		
QY	181	AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr	200		
Db	609	GCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGACA	668		
QY	201	ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg	220		
Db	669	CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAAAATCAGAGAGA	728		
QY	221	ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg	240		
Db	729	ACACTGGAATGTTACGTTTCATAACTACTAAGGATTAGTGTATATTTTCCAAACCTTGAGG	788		
QY	241	HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg	260		
Db	789	CATGAATTTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGGATGTGAATGCATCCCGG	848		
QY	261	GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr	280		
Db	849	CAGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCACG	908		
QY	281	GluGlyLeuPheAsnMetAspGluuaspGluGluThrGluHisGluThrLysAlaGlyPro	300		
Db	909	GAAGGATTGTTTATATATGATGAAGATGAAGAACTGAACATGAACAAGGCTGGTCTCT	968		
QY	301	GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu	320		
Db	969	GAACGGCTCGACCGATGGTGCATCTCTGTAGCCGAGCGCTGGACATCTCTGATGTCTTTG	1028		
QY	321	ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr	340		
Db	1029	GTTTGTCTCTACATGAAGGATGTCTGCTATGTAGATGGTAAGTTGATAACGGCAAAACA	1088		
QY	341	LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla	360		
Db	1089	AAGGATCTATATCCGACCTGTGATAAACATCTTTGACAAACTCTCTGTGCCCCACCATGCC	1148		
QY	361	SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu	380		
Db	1149	TCCTGCCATGTACAGTTTTTTCATGTTTTACCTCTGTAGTTTCAAATGGGATTCCGAGAG	1208		
QY	381	AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg	400		
Db	1209	GCATTTTGGAACATCTCTGGAAAAAATTTGCAGGACCCCAAGTAATCTCTGCCATCATCAGG	1268		
QY	401	GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle	420		
Db	1269	CAGGCTGCTGGAATATATATGGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATT	1328		
QY	421	ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln	440		



Db 1329 ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTTAATAACCAG 1388

QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPropheTyrSerAlaCys 460  
|||||

Db 1389 GATTCGGGAACAAAGGCATTCTCGGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC 1448

QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
|||||

Db 1449 CAAGCTGTGTTCTACACCTTTGTTTATAGACACAAGCAGCTTTTGAGCGGAACCTGAAA 1508

QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
|||||

Db 1509 GAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1568

QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520  
|||||

Db 1569 CCCCTGAAGATTGCCTGCCCTCAGTGGTTAACTTTTGTGCTGCAATCACAAATAAGTAC 1628

QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540  
|||||

Db 1629 CAGCTCGTCTTCTGCTACACCATCATTTGAGAGAACAAATCGCCAGATGCTGCCAGTCAIT 1688

QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnPheLeuAspThrPhePhe 560  
|||||

Db 1689 AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCGCTGGACACCTTCTTC 1748

QY 561 PropheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
|||||

Db 1749 CCCTTTGATCCCTGTGCTGAAGAGGTCAAGAAATTCATTGATCCTATTATCAGGTA 1808

QY 581 TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
|||||

Db 1809 TGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAGGACATA 1868

QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
|||||

Db 1869 GTGGAAGATGAAGATGATGACTTCTGAAAGCGAAGTGCCCCAGAAATGATACCGTGATT 1928

QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer 640  
|||||

Db 1929 GGGATCACACCAAGCTCCTTTGACACGCATTCCGAAGTCTTCAAGTAGTGTGGGCTCC 1988

QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
|||||

Db 1989 CCACCCGTGTTGTACATGCAACCCAGTCCCCCTC 2021

RESULT 2

HSA272050 3756 bp mRNA linear PRI 01-AUG-2002

LOCUS

DEFINITION Homo sapiens mRNA for transcription initiation factor IA protein (TIF-IA gene).

ACCESSION AJ272050

VERSION AJ272050.1 GI:10046713

KEYWORDS TIF-IA; transcription initiation factor.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Bodem,J., Dobрева,G., Hoffmann-Rohrer,U., Iben,S., Zentgraf,H., Delius,H., Vingron,M. and Grummt,I.

TITLE TIF-IA, the factor mediating growth-dependent control of ribosomal RNA synthesis, is the mammalian homolog of yeast Rrn3p

JOURNAL EMBO Rep. 1 (2), 171-175 (2000)

MEDLINE 21163499

PUBMED 11265758

REFERENCE 2 (bases 1 to 3756)

AUTHORS Bodem,J.J.

Direct Submission

TITLE Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg, GERMANY

FEATURES Location/Qualifiers

source 1. .3756  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
gene 1. .3756  
/gene="TIF-IA"  
23. .1978  
/gene="TIF-IA"  
/function="transcription initiation factor"  
/codon\_start=1  
/evidence=experimental  
/product="transcription initiation factor IA protein"  
/protein\_id="CAC07955.1"  
/db\_xref="GI:10046714"  
/db\_xref="GOA:Q9H4F0"  
/db\_xref="SPTREMBL:Q9H4F0"  
/translation="MAAPLLHTRLPDAAASSAVKKLGASRTGISNMRALENDFFNS  
PPRKTVRFGGTVTEVLLKYKKGTRNDPELLKNQLDPPDIKDDQIINWLLFRSSIMYL  
TKDFEQLISIIILPWLNRSQTVVEEYLAFLGNLSAQTVFLRPLCLSMIASHFVPRV  
IIKEGDVDVSDDEDDNLPAFDTCHRALQIIARYVPSTPWFLMPILVEKFFPVRKS  
ERTLECYVHNLRIISVYFPTLRHEILELIIIEKLKLDVNASRQGIEDAETATQTCGG  
TDSTEGLFNMEDEETEHETKAGERLDQMVPVPAERLDILMSLVLSYMKDVCYVDGK  
VNGKTKLDLYRDLINIFDKLLPHTASCHVQFMFYLCSFKLGFABAFLEHLWKQLQD  
PSNPARIROAAGNYIGSFLARAKFIPLITVSKCLDLLVNLHIYLNNDQSGTKAFCDV  
ALHGPFIYACQAVFYXVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS  
VWNFFAAITNKYQLVFCYTIERNRQMLPVIRSTAGDSVQICTNPLDTFFFPDPCV  
LKRSKKFIPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDFLKGEVPPQNDTVIGITP  
SSFDTHFRSPSSSVGSPVLYMQPSPL"  
1979. .3756  
/gene="TIF-IA"  
/evidence=experimental

3'UTR

ORIGIN

Alignment Scores:

Pred. No.: 1.37e-315 Length: 3756

Score: 3398.00 Matches: 650

Percent Similarity: 99.85% Conservative: 0

Best Local Similarity: 99.85% Mismatches: 1

Query Match: 99.85% Indels: 0

DB: 9 Gaps: 0

US-09-932-678-2 (1-651) x HSA272050 (1-3756)

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAalaSerSerSerAla 20  
|||||

Db 23 ATGGCGGCACCGCTGCTTCACACGCGTTTCCCGGGAGATGCGGCCGCTTCCTCTGCA 82  
|||||

QY 21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuLeuAsnAsp 40  
|||||

Db 83 GTTAAGAAAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATCGTGATTAGAGAATGAC 142  
|||||

QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
|||||

Db 143 TTTTTCATTTCTCCCCAAGAAAACTGTTGCGTTTGGTGGAACTGTGACAGAAGTCTTG 202  
|||||

QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
|||||

Db 203 CTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCACTGTTAGAT 262  
|||||

QY 81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
|||||

Db 263 CCAGACATAAAGGATGACCAGATCATCACTGGCTGCTAGAAATTCGGTCTTCTATCATG 322  
|||||

QY 101 TyrLeuThrLysAspPheGluGlnLeuLeuSerIleIleLeuArgLeuProTrpLeuAsn 120  
|||||

Db 323 TACTTGACAAAAGACTTTGAGCAACTTATCATGATTATATTAGATTGCTTGGTTGAAT 382  
|||||

QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
|||||

Db 383 AGAAGTCAAACAGTAGTGAAGAGATATTGGCTTTTCTGGTAAATCTTGTATCAGCACAG 442  
|||||

QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
|||||

Db 443 ACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGGCTCCCCGAGTG 502

Qy 161 ILeIleLySGLuGLyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180

Db 503 ATCATTAAGAAAGCGGATGATGATGTTTCAGATTCTGATGATGAAGATATAATCTTCCT 562

Qy 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200

Db 563 GCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGACA 622

Qy 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220

Db 623 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTCGAAAAATCAGAGAGA 682

Qy 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240

Db 683 ACACGTGGAATGTTACGTTTCATAACTTACTAAGGATTAGTGTATATTTTCCAAACCTTGAGG 742

Qy 241 HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260

Db 743 CATGAAATTCCTGGAGCTTATTATTGAAAACTACTCAAGTTGGATGTGAATGCATCCCCG 802

Qy 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280

Db 803 CAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTTGGTGGGACAGATTCCACG 862

Qy 281 GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro 300

Db 863 GAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCTCT 922

Qy 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320

Db 923 GAACGGCTCGACACAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCCTGATGCTTTG 982

Qy 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340

Db 983 GTTTTGTCTACATGAAGGATGTCTGTATGATGGTAAGGTTGATAACGGCAAAACA 1042

Qy 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360

Db 1043 AAGGATCTATATCGGACCTGATAAACATCTTTGACAAACTCTGTTGCCACCCATGCC 1102

Qy 361 SerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380

Db 1103 TCCTGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTAGTTTCAAATTTGGGATTCGACAG 1162

Qy 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400

Db 1163 GCATTTTGGAAACATCTCTGAAAAAAATTCAGAGACCCCACTAAATCTCTGCCATCATCAGG 1222

Qy 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420

Db 1223 CAGGCTGCTGGAATAATTATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCTCTTATT 1282

Qy 421 ThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440

Db 1283 ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAG 1342

Qy 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460

Db 1343 GATTCCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC 1402

Qy 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480

Db 1403 CAAGCTGTGTTCTACRCCTTTGTTTTTAGACAACAAGCAGCTTTTGGCGGAAACCTGAAA 1462

Qy 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500

Db 1463 GAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGGCGGATAGTGATGAGCCAGCTAAAT 1522

Qy 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr 520

Db 1523 CCCCTGAAGATTGCTGCCCTCAGTGGTTAACTTTTTTGTGCTGCAATCACAAATAAGTAC 1582

Qy 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540

Db 1583 CAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCATT 1642

Qy 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560

Db 1643 AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTC 1702

Qy 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580

Db 1703 CCTTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTG 1762

Qy 581 TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600

Db 1763 TGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 1822

Qy 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620

Db 1823 GTGGAAGATGAAGATGATGACTTTTCTGAAAGGCGAAGTGCCCAAGATGATACCGTGATT 1882

Qy 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer 640

Db 1883 GGGATCACACCAAGTCTCTTTGACACGCAATTCGGAAGTCTTCAAGTAGTGTGGGCTCC 1942

Qy 641 ProProValLeuTyrMetGlnProSerProLeu 651

Db 1943 CCACCCGTGTTGTACATGCAACCCAGTCCCTC 1975

RESULT 3

AX036047

LOCUS AX036047 2040 bp DNA linear PAT 15-NOV-2000

DEFINITION Sequence 1 from Patent WO0055316.

ACCESSION AX036047

VERSION AX036047.1 GI:11191586

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Grummt,I. and Vingron,M.

TITLE Rna polymerase i transcription factor tif-ia

JOURNAL Patent: WO 0055316-A 1 21-SEP-2000; DEUTSCHES KREBSFORSCH (DE) ; GRUMMT INGRID (DE) ; VINGRON MARTIN (DE)

FEATURES

source Location/Qualifiers

1. .2040

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

40. .1995

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC16268.1"

/db\_xref="GI:11191587"

/db\_xref="REMTREMBL:CAC16268"

/translation="MAAPLLHTPELPGDAAASSSAVKKLGASRTGISNMRALENDFFNSPPRKTVRFGGTVTEVLKYYKKGTEINDFELLKNQLLDDPDIKDDQIINWLLFRSSIMYLTKDFEQLSIIILRLPWLNRSQTVVEEYLAFLGNLVSAQTVELRPLCLSMIASHFVPRVITKEGDVDSDDDDNLPANPDTCHRALQIIARIYVPSTPWFPLMPILVEKFPFVRKSERTLECYVHXLLRISVYFPTLRHEILELIEKLLKLDVNASRQGTEDAEETATQTCGGTDSTEGLFNMDDEETETHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGKVDNGKTKDLYRDLINIFDKLLPHTASHCVQFMFYLCSFKLGFAEAFLEHLWKLIQDPSNPALIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNLWHIYLNQDSGTGKAFCDVALHGPFYSACQAVFYXFVERHKQLLSGNLKEGLQYLSLNFERIVMSQLNPLKICLPSVYNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDFTFFFPDPCVLKRSKKFIDPIYQWEDMSAEELQEPKKPMKKDIVEDEDDDFLKGEVFPQNDTVIGITPSSFDTHFRSPSSSVGSPPVLYMQPSPL"

ORIGIN

Alignment Scores: 2.7e-315 Length: 2040

Pred. No.:

Score:		3391.00	Matches:	649
Percent Similarity:	99.69%		Conservative:	0
Best Local Similarity:	99.69%		Mismatches:	2
Query Match:	99.65%		Indels:	0
DB:	6		Gaps:	0
US-09-932-678-2 (1-651) x AX036047 (1-2040)				
QY	1	MetAlaAProLeuLeuHisThrArgLeuProGlyAspAlaAlaAAserSerAla	20	
Db	40	ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGAGATCGCGCGCTTCGTCCTCTGCA	99	
QY	21	VallLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp	40	
Db	100	GTTAAGAAGCTGGCGCGTCGAGACTGGGATTTCAAATATGCGTGCAATTAGAGAAATGAC	159	
QY	41	PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu	60	
Db	160	TTTTTCAATTCTCCCCAAGAAAAAAGTTCGGTTTGGTGGAATGTGACAGAGTCTTG	219	
QY	61	LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuAsp	80	
Db	220	CTGAAGTACAAAAAGGTGAACAAATGACTTGGATTGTTGAAGAACCCAGCTGTTAGAT	279	
QY	81	ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet	100	
Db	280	CCAGACATAAAGGATGACCAGATCATCAACTGGCTGTAGAAATCCGTTCTTCTATCATG	339	
QY	101	TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn	120	
Db	340	TACTTGACAAAAAGACTTTTGAGCAACTTATCAGTATTATATTAAGATTGGCTTGGTTGAAT	399	
QY	121	ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln	140	
Db	400	AGAAGTCAAAACAGTAGTGAAGAGTATTGGCTTTTCTTGGTAAATCTTGATCAGCACAG	459	
QY	141	ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal	160	
Db	460	ACTGTTTTCTCAGACCGTGCTCAGCATGATGCTTCCCATTTTGTGCTCCCCGAGTG	519	
QY	161	IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro	180	
Db	520	ATCATTAAGGAAGCGATGTAGATGTTTTCAGATTCTGATGATGAAGATGATAATCTTCCT	579	
QY	181	AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr	200	
Db	580	GCAAAATTTGACACATGTCAAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGACA	639	
QY	201	ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg	220	
Db	640	CCGTGGTTTCTCATGCCAATACTGGTGGAATAATTTCCATTTGTTGGAATAATCAGAGAGA	699	
QY	221	ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg	240	
Db	700	ACACTGGAATGTTACGTTTCATNACTTACTAAGGATTAGTGATATATTTTCCAAACCTTGAGG	759	
QY	241	HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg	260	
Db	760	CATGAAATTTCTGGAGCTTATTATTGAAAACTACTCAAGTTGGATGTGAATGCATCCCCG	819	
QY	261	GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr	280	
Db	820	CAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGACAGATTCCACG	879	
QY	281	GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro	300	
Db	880	GAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCT	939	
QY	301	GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu	320	
Db	940	GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCCTTG	999	
QY	321	ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr	340	

Db	1000	GTTTTGTCTACATGAAGGATGCTCTGCTATGTAGATGGTAAGGTGATAACGGCAAAACA	1059	
QY	341	LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla	360	
Db	1060	AAGGATCTATATCGGACCTGATAAAACATCTTTTGACAAACTCCTGTTGCCCAACCATGCC	1119	
QY	361	SerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu	380	
Db	1120	TCCTGCCATGTACAGTTTTTTCATGTTTTACCTCTGTAGTTTCAAATTTGGGATTTCGAGAG	1179	
QY	381	AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg	400	
Db	1180	GCATTTTGGAAACATCTCTGGAAAAAATTGCAGGACCCCAAGTAATCCTGCCATCATCAGG	1239	
QY	401	GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle	420	
Db	1240	CAGGCTGCTGGAAATTATATTGAAGCTTTTGGCAAGAGCTAAATTTTATTCCTCTTATT	1299	
QY	421	ThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln	440	
Db	1300	ACTGTAATAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA	1359	
QY	441	AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys	460	
Db	1360	GATTCGGGAACAAAGGCAATCTCGCATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC	1419	
QY	461	GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys	480	
Db	1420	CAAGCTGTGTTCTACNCCTTTGTTTTAGACACAAAGCAGCTTTTGACGGGAACCTGAAA	1479	
QY	481	GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn	500	
Db	1480	GAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGACGGGATAGTGATGAGCCAGCTAAAT	1539	
QY	501	ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr	520	
Db	1540	CCCCCTGAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTGTGCAATCACAAATAAGTAC	1599	
QY	521	GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle	540	
Db	1600	CAGCTCGTCTTCTGTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCAAT	1659	
QY	541	ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe	560	
Db	1660	AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCCGCTGGACACCTTCTTC	1719	
QY	561	ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal	580	
Db	1720	CCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTG	1779	
QY	581	TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle	600	
Db	1780	TGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCCATGAAAAAGGACATA	1839	
QY	601	ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle	620	
Db	1840	GTGGAAGATGAAGATGATGACTTTCTGAAAGCGGAAGTCCCGCCAGAAATGATACCGTGATT	1899	
QY	621	GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer	640	
Db	1900	GGGATCACACCAAGCTCCTTTGACACGCAATTTCCGAAAGTCCTTCAAGTAGTGTGGGCTCC	1959	
QY	641	ProProValLeuTyrMetGlnProSerProLeu	651	
Db	1960	CCACCCCGTGTGTACATGCACCCAGTCCCTC	1992	

RESULT 4  
BC036182  
LOCUS  
DEFINITION Homo sapiens RNA polymerase I transcription factor RRN3, mRNA (cdna clone IMAGE:5298581), partial cds.  
ACCESSION BC036182

BC036182.1 GI:23243332

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens (human)  
Homo sapiens

REFERENCE  
AUTHORS

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, A.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.M., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE

22388257

PUBMED

12477932

REFERENCE

2 (bases 1 to 3709)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (31-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 63 Row: 0 Column: 6  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361630.  
Location/Qualifiers  
1. .3709  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5298581"  
/tissue\_type="Testis"  
/clone\_lib="NIH MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

FEATURES  
source

6.29e-315 Length: 3709  
3391.00 Matches: 649  
99.85% Conservativeness: 1  
99.69% Mismatches: 1

ALIGNMENT  
Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:

Query Match: 99.65% Indels: 0  
DB: 9 Gaps: 0  
US-09-932-678-2 (1-651) x BC036182 (1-3709)

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAalaSerSerSerAla 20  
Db 18 ATGGCGGCACCGCTGCTTCACACGCGTTTGGCGGAGATCGCGCGCTTCGCTCTGCA 77  
QY 21 VallysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
Db 78 GTTAAAGAGCTGGCGCTCGAGGACTGGGATTTCAAATATGGTGCTTAGAATGAC 137  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
Db 138 TTTTTCATTTCTCCCCCAAGAAAAAAGTTCGTTGGTGGTGGTGGTGGTGGTGGTGGT 197  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
Db 198 CTGAAGTACAAAAAGGTTGAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTAGAT 257  
QY 81 ProAspIleLysAspAspGlnIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
Db 258 CCAGACATAAAGGATGACCATCATCACTGCTGCTAGAAATTCGTTCTTCTATCATG 317  
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleLeuArgLeuProTrpLeuAsn 120  
Db 318 TACTTGACAAAAGACTTTGAGCAACTTATCATGATATATTAAGATTGCCTTGGTGAAT 377  
QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
Db 378 AGAAGTCAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAG 437  
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProArgVal 160  
Db 438 ACTGTTTTCCTCAGACCGTGTCTCAGCATGATGCTTCCCATTTGTGCTTCCGAGTG 497  
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180  
Db 498 ATCATTAAGGAAGCGGATGTAGATGTTTTCAGATTTCTGATGATGAAGATGATATCTCCT 557  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgTyrValProSerThr 200  
Db 558 GCAATTTTGCACATGTCTCAGAGCGCTTGCAATAATAGCAAGATATGTACCATCGACA 617  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
Db 618 CCGTGGTTTCTCATGCCAATACTGTTGGTGGAAAAATTTCCATTTGTTGAAAAATCAGAGA 677  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
Db 678 ACATGGATGTTACGTTTCATAACTTACTAAGGATAGTGTATATTTTCCAACTTGAGG 737  
QY 241 HisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
Db 738 CATGAAATTTCTGGAGCTTATTATTGAAAAAATCTACTCAAGTTGGATGTGAATGCCATCCC 797  
QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
Db 798 CAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTTGTTGGTGGGACAGATCCACG 857  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
Db 858 GAAGGATTGTTTAAATATGGATGAAGATGAAGAACTGAACATGAACAAAGGCTGTCCT 917  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
Db 918 GAACGGCTCGACCATGATGGTGCATCTCTGTAGCGAGCGCTGGACATCTCTGATCTTTG 977  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
Db 978 GTTTTGTCTACATGAAGGATGTCTGCTATGTAGATGGTGAAGGTGATAACGGCGAAACA 1037

QY	341	LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla	360
Db	1038	AAGGATCTATATCGGACCTGTATAAACATCTTTGACAAACTCTGTGTGCCACCCATGCC	1097
QY	361	SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu	380
Db	1098	TCCTGCCATGACAGTTTTTCATGTTTTACCTCTGTAGTTTCAAAATGGGAATTCGCAGAG	1157
QY	381	AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg	400
Db	1158	GCATTTTGGAAACATCTCTGGAAAAAATTGCAGGACCCCAAGTAATCTCGCCATCATCAGG	1217
QY	401	GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle	420
Db	1218	CAGGCTGCTGGAATTTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATT	1277
QY	421	ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln	440
Db	1278	ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAAATAACCAG	1337
QY	441	AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys	460
Db	1338	GATTCGGGAACAAAGGCATCTCGCATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC	1397
QY	461	GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys	480
Db	1398	CAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGCGGAACCTGAAA	1457
QY	481	GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn	500
Db	1458	GAAGTTTTCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT	1517
QY	501	ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr	520
Db	1518	CCCCTGAAGATTGCCCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAAATAAGTAC	1577
QY	521	GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle	540
Db	1578	CAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAAATGCCAGATGTCGCCAGTCATT	1637
QY	541	ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe	560
Db	1638	AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTC	1697
QY	561	ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal	580
Db	1698	CCCTTTGATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTG	1757
QY	581	TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle	600
Db	1758	TGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAGAAACCCCATGAAAAGGACATA	1817
QY	601	ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle	620
Db	1818	GTGGAAGATGAAGATGATGACTTCTGAAAGCGGAAGTGCCCCAGAAATGATACCGTGATT	1877
QY	621	GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer	640
Db	1878	GGGATCACACCAAGACTCCTTTGACACGCATTTCCGAAGTCCCTCAAGTAGTGGGCTCC	1937
QY	641	ProProValLeuTyrMetGlnProSerProLeu	651
Db	1938	CCACCCGTGTTGTACATGCAACCCAGTCCCCCTC	1970
RESULT	5		
LOCUS	BC034110		
DEFINITION	Mus musculus expressed sequence AL023001, mRNA (cdna clone IMAGE:4504751), partial cds.		
ACCESSION	BC034110		
VERSION	BC034110.1	GI:21706612	
KEYWORDS	.		
SOURCE	Mus musculus (house mouse)		

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 3166)
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 3166)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: The Cepko Laboratory cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www-shgc.stanford.edu">http://www-shgc.stanford.edu</a> Contact: (Dickson, Mark) <a href="mailto:mcd@paxil.stanford.edu">mcd@paxil.stanford.edu</a> Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
FEATURES	Location/Qualifiers
source	1..3166 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:4504751" /tissue_type="Eye, retina, mouse strain C57Bl\6" /clone_lib="NIH MGC_94" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" <1..3166 /gene="AL023001" /db_xref="LocusID:106298" /db_xref="MGI:2146525" <1..1544 /gene="AL023001" /codon_start=3 /product="AL023001 protein" /protein_id="AAH34110.1" /db_xref="GI:21706613"
gene	
CDS	

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 44 Row: g Column: 5  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.



/db\_xref="LocusID:106298"  
/db\_xref="MGI:2146525"  
/translation="PCLSMIASHFVPPRVIVKEGGIDVSDSDDEDDNLPAIFDTCHRA  
LQIITRYVSPWFMPILVVEKFPFVRKSERLTCEYVHNLRLISLYFPTLRRIILELV  
IEKLLKLDVSVSRQDIEDAEERAAQTCCGTDTEGLFNWDEDDTDPEKKADQBPNQ  
MAHPTAERLDVLLCLLSYIEDVCRVHGKIDNNKTOKLYDLISIFDKLLPLTHASCH  
VQFEMFFLCSFKLGFAEAFLEHLWKKLQDPNPAIRQAAANYIGSFLARAKFIPLIT  
VKTCLDLLVNLMHMYLTNQDSGTFKAFCDVALHGPFYSACQAVYTVVFRHKQLLSGNL  
KQGLQYLQSLNFERIVLSQNLPLKICLPQVNVFFAAITNKYQLVFCYTIMERNRQML  
PWRSTAGGDSVQCTNPLDTFPFDPCLVKRSKKFIDPIYQIWEDGSAEELQEFKKS  
TKKEVVEDEDDFLKGEVPQSDTVTGLTPSSFDTHFQSPSSSVGSPPLVIYFQSPPLL  
TRYD"

ORIGIN

Alignment Scores:  
Pred. No.: 2,87e-217 Length: 3166  
Score: 2372.50 Matches: 444  
Percent Similarity: 93.89% Conservative: 32  
Best Local Similarity: 87.57% Mismatches: 30  
Query Match: 69.72% Indels: 1  
DB: 10 Gaps: 1

US-09-932-678-2 (1-651) x BC034110 (1-3166)

QY	146	ProCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleIleLysGluGly	165
Db	3	CCATGTCTCAGCATGATTGGCTCTCATTTTGTACCTCCCGAGTAATTGTCAAGGAAGGT	62
QY	166	AspValAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThr	185
Db	63	GGCATAGATGTTTCAGATTCTGATGACGAAGATGACAACTTCCTGCAATTTTGACACA	122
QY	186	CysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMet	205
Db	123	TGTCACAGAGCCTTGCAATAATAACAAGATATGTCCCATCGACACCATGGTTCTAATG	182
QY	206	ProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyr	225
Db	183	CCAATACTGGTGGAGAAATTCGCGTTTGTGAGGAAGTCCGAGAGAACATTGGAATGTTAT	242
QY	226	ValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGlu	245
Db	243	GTTCATAACTTATTAGGATAAGTTTATATTCCCAACTTTGAGGCGTGAATTTCTGGAG	302
QY	246	LeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAsp	265
Db	303	CTTGTTATTGAAAAGCTACTCAAGTTAGATGTGAGTGATGTCGGGCAGGATATTGAAGAT	362
QY	266	AlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn	285
Db	363	GCTGAAGAGAAAGCAGCTCAGACTTGTGTGGGACAGATACCACAGAAGGACTGTTTAAT	422
QY	286	MetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGln	305
Db	423	ATGGATGAAGATGAGGACACTGACCCTGAGAAGAAAGCTGACCAGGAGCAGCCTAACCA	482
QY	306	MetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMet	325
Db	483	ATGGCTCATCCACCGCAGAGCGCCTGGACGTCCTGCTGTGCTTGTCTGTCTCTACATT	542
QY	326	LysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArg	345
Db	543	GAGGATGTCGCCGTGTGCACGGTAAATTGATAACAATAAAACAAGGATTATACCGT	602
QY	346	AspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGln	365
Db	603	GATCTGATATCCATCTTTTGACAAACTTCTGTGCCCCACACATGCCTCCTGCCATGTACAG	662
QY	366	PhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis	385
Db	663	TTCTTCATGTTTTCCTCTCGAGCTTCAAGTTGGGATTTCAGAGCAATTTTGGAACAT	722
QY	386	LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn	405

Db	723	CTCTGGAAAAAGTTGCAGGATCCAATAAACCCCGCCCATCATCAGGCAGGCTGCTGCAAA	782
QY	406	TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys	425
Db	783	TATATTGGGAGCTTTTGGCCAGAGCTAAATTTATTCTCTCTTATCACTGTGAAGACATGC	842
QY	426	LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys	445
Db	843	CTGGATCTCTTGTTAACTGGCTGCACATGTACCTTACTTAACCAAGGATTCGGGAACAAAG	902
QY	446	AlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyr	465
Db	903	GCTTTTGTGACGTTGCACTCCATGGACCATTTTATTAGCCTGCCAAGCTGTGTTCTAC	962
QY	466	ThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr	485
Db	963	ACTGTTGTTTGTAGACACAAGCAGCTTTTGAGTGGAACTTGAAGCAAGGTCTACAGTAT	1022
QY	486	LeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCys	505
Db	1023	CTTCAGAGTCTAAATTTTGAGCGCATTTGTGTGAGCCAGCTGAACCCACTGAAGATCTGC	1082
QY	506	LeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCys	525
Db	1083	CTGCCGCAAGTGGTTAATTTCTTTGTGCTGCTATCACAAATAAATACCAGTGGTGTCTGC	1142
QY	526	TyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGly	545
Db	1143	TACACCATCATGGAGAGAAACAGTCGCCAGATGCTCCCGTATCCGCAGCACCGCTGGT	1202
QY	546	GlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCys	565
Db	1203	GGAGACTCCGTCAGACCTGCACCAACCCACTGGACACTTCTTCCCTTTGACCCCTTGT	1262
QY	566	ValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspMetSer	585
Db	1263	GTGCTTAAGAGGTCAAAGAAAGTTCAATTCATTCATCTATTATCAGATTTGGGAAGATGGGAGT	1322
QY	586	AlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAsp	605
Db	1323	GCTGAAGAGCTTCAGGAGTTTAAGAAATCTACTAAAAAGGAGGTAGTGGAGGATGAAGAT	1382
QY	606	AspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer	625
Db	1383	GATGACTTTTGAAGGCGAGGTGCCCCAGAGTGACACAGTGAAGTGGCCTTACTCCGAGC	1442
QY	626	SerPheAspThrHisPheArgSerProSerSerSerValGlySerProProValLeuTyr	645
Db	1443	TCCTTTGATACCCACTTCCAAAAGTCCTCCAGTAGTGTGGGCTCCCTCCTGTGCTGTAT	1502
QY	646	MetGln--ProSerProLeu 651	
Db	1503	ATACCAAGGCCAGTCTCCACTC 1523	
RESULT 6			
AC012547			
LOCUS			
DEFINITION			
AC012547			
AC012547.2			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			



Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funk,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (29-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced gi:6139119.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L3178  
Center clone name: ll\_K\_1  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 127596 bases at least Q40  
Consensus quality: 130929 bases at least Q30  
Consensus quality: 132388 bases at least Q20  
Insert size: 133000; agarose-fp  
Quality coverage: 4.9 in Q20 bases; agarose-fp  
Quality coverage: 4.9 in Q20 bases; sum-of-contigs  
-----

TITLE  
JOURNAL  
COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1516: contig of 1516 bp in length  
\* 1517 1616: gap of 100 bp  
\* 1617 3136: contig of 1520 bp in length  
\* 3137 3236: gap of 100 bp  
\* 3237 5148: contig of 1912 bp in length  
\* 5149 5248: gap of 100 bp  
\* 5249 7742: contig of 2494 bp in length  
\* 7743 7842: gap of 100 bp  
\* 7843 10650: contig of 2808 bp in length  
\* 10651 10750: gap of 100 bp  
\* 10751 16194: contig of 5444 bp in length  
\* 16195 16294: gap of 100 bp  
\* 16295 25751: contig of 9457 bp in length  
\* 25752 25851: gap of 100 bp  
\* 25852 37209: contig of 11358 bp in length  
\* 37210 37309: gap of 100 bp  
\* 37310 47828: contig of 10519 bp in length  
\* 47829 47928: gap of 100 bp  
\* 47929 64239: contig of 16311 bp in length  
\* 64240 64339: gap of 100 bp  
\* 64340 77846: contig of 13507 bp in length  
\* 77847 77946: gap of 100 bp  
\* 77947 95311: contig of 17365 bp in length  
\* 95312 95411: gap of 100 bp  
\* 95412 108930: contig of 13519 bp in length

\* 108931 109030: gap of 100 bp  
\* 109031 135873: contig of 26843 bp in length.  
FEATURES  
source  
1. .135873  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-11K1"  
/clone\_lib="RPCI-11 Human Male BAC"  
1. .1516  
/note="assembly\_fragment"  
1617. .3136  
/note="assembly\_fragment"  
3237. .5148  
/note="assembly\_fragment"  
5249. .7742  
/note="assembly\_fragment"  
7843. .10650  
/note="assembly\_fragment"  
10751. .16194  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right"  
16295. .25751  
/note="assembly\_fragment"  
25852. .37209  
/note="assembly\_fragment"  
37310. .47828  
/note="assembly\_fragment"  
47929. .64239  
/note="assembly\_fragment"  
64340. .77846  
/note="assembly\_fragment"  
77947. .95311  
/note="assembly\_fragment"  
95412. .108930  
/note="assembly\_fragment"  
109031. .135873  
/note="assembly\_fragment"  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.89e-183 Length: 135873  
Score: 2039.00 Matches: 453  
Percent Similarity: 77.71% Conservative: 56  
Best Local Similarity: 69.16% Mismatches: 120  
Query Match: 59.92% Indels: 32  
DB: 2 Gaps: 13  
US-09-932-678-2 (1-651) x AC012547 (1-135873)  
QY 2 AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAalaSerSerAlaVal 21  
Db 96743 GCAACCAGGCTGCTTCTCACATGGTTGCCGGGAA-GCAGAGGCTTCAGCTTCTGTGTC 96801  
QY 22 LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe 41  
Db 96802 AAGAAGCTGGGTATGTCATGGACCAGTATTTCAGATATGCTGCATTCAAGATGATTTT 96861  
QY 42 PheAsnSerProArgLysThrValArgPheGlyThrValThrGluValLeuLeu 61  
Db 96862 TCCAGCTCTCCCCAAGAAAAATTGTTTCATCTTGATGGAACGGGACAGAACTCTTGCCG 96921  
QY 62 LysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspPro 81  
Db 96922 AAGACAAAGAGTGAAACAAATTAACTTTGAGCTGTTGAAGAGCCAGTTGTGAGATTCA 96981  
QY 82 AspIleLysAspAspGlnIlelleAsnTrpLeuLeuGluPheArgSerSerIleMetTyr 101  
Db 96982 AACATAAGGATGACCAGATCTTGAACCTGGTGTGAAATTTTGCTCTTCTATCATGTAC 97041  
QY 102 LeuThrLysAspPheGluGlnLeuIleSerIlelleLeuArgLeuProTrpLeuAsnArg 121

Db 97042 TTGGCAAAAGACTTTTGAGCAGCTCCTCAGTATTACATTGAGATTGCCTTGGTTGAATAGA 97101

QY 122 SerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThr 141

Db 97102 ATTCAAACAGTGGTGAAGAGTATTTTGGCTTCTCTGGGTAGTCTTTGTAAACAAC-AGA--- 97157

QY 142 ValPheLeuArgProCysLeuSerMetIleAlaSerHisPheVal---ProProArgVal 160

Db 97158 CTCTTCCTTAGCCCCAAGTCTCAGCATAATTGCTTCCCATTTGGCGTACCCCTCCTTGAATG 97217

QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180

Db 97218 ATCATTAAGAAGAGCCCATGTAAATGTTTCAGATTCTGTATGATGGAGATGATTATCTTCCT 97277

QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200

Db 97278 GCAGATTTTGACACACGTCATACAACCTTACAATAACAGCAAGATATGTCCCACTGGCA 97337

QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220

Db 97338 CTGTGTTTTCTTTGCCCATACTGGTGAACAACATTTCCATTGTT---CAATCAGAGAGA 97394

QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240

Db 97395 ACACCTGGAATGTTACATTCTATAACTTACTAAAGATTAGTATATATCTTTCCAACTTGAGG 97454

QY 241 HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260

Db 97455 AATGAAATCTGAAGCTTATTATTAAAAATCTATTCAAATGGATGTGAATACATTCTCG 97514

QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280

Db 97515 CATGATATTGAAGATGCTGAGGAACACAGCAGCTCAAACCTGTGGTGGGACAGATTCCATG 97574

QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300

Db 97575 GAATGATTGTTCAATATGAACGAAGATGAAGAGACT-----GTTTCTAAGGCTGACTCC 97628

QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320

Db 97629 AAATGGCTCGGTCAGATGGTGCACCCCTGTAGCTGAATGCTGGACACCCCTGTCTTTG 97688

QY 321 ValLeuSerTyrMetLysAspValCysTyrValAsp-----GlyLysValAspAsnGly 338

Db 97689 CTTTTGCAGTACTTGAAGTATGTCTGCTATGTAAA-ATACATGGTAAGGTGATAACAAC 97747

QY 339 LysThrLysAspLeu-----TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeu 356

Db 97748 AAAATAAAGGATTT-ATAATATATCACAACTGTATAACATATTTGACAAA----- 97797

QY 357 ProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeu 376

Db 97798 ---CCCCCT-GCCTCCTGTCACGTACAGTTTTTATGTTTTACCTCTGTGGCTTAAAATG 97853

QY 377 GlyPheAlaGluAlaPheLeuGluHisLeuTrp---LysLysLeuGlnAspProSerAsn 395

Db 97854 CGACTCTCAGAAGCATTTTGGAACGTCCTCTGGGAAAAAAAATGCAGGATCCAAATAAT 97913

QY 396 ProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLys 415

Db 97914 CCTGTCTCATCAGGTGAGTGTCTGGAATATTATATGGAAGCTCTTTGGCAAAAGCTAAA 97973

QY 416 PheIleProLeuIle---ThrValLysSerCysLeuAspLeuValAsnTrpLeuHis 434

Db 97974 TTCATTCTCTCATTAACAACTATAAAATCATGCCCTAGGCTTTTCGGTTAACTGGCCACAC 98033

QY 435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454

Db 98034 AAATACCTTACTAACCCAGGATCTGAGAATGAAGGCTTTTGTGATATTGCTCTCCATCAA 98093

QY 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474

Db 98094 ACATTTTACTCAGCCTTCCAAGCTGTGTTCCACACCTTTGTTTTTTAGACACAAGCAACTT 98153

QY 475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494

Db 98154 TTAAGCGTAAACCTGAAAGAAGGTTTGAGATGTCTTCAGTGCCTACATTTTAAGGGATA 98213

QY 495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla 514

Db 98214 GTGACTAGGCAGCTAAATCCTCTCTGAATATTTGCCTGCCCTCAGTGGTTAACTTATTGCT 98273

QY 515 AlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn-Ar 534

Db 98274 GCTGTACGAATAAATATCAGCTAGTC---TGCTACACTATCATTTGAGAGGAACAACGGC 98330

QY 534 gGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAs 554

Db 98331 C-----AGAAATACAGCTGGAGGAGACTCGGTGTGAACGTGCACCAA 98372

QY 554 nProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPheIl 574

Db 98373 CCCATTGGACGCCCTTCITCCCTTTGATCCTTGTGCGCTTAAGCAGTCAAGGAAAGTCAT 98432

QY 574 eAspProIleTyrGlnValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLy 594

Db 98433 TGATTCTATTATCAGATATGGGAAGACATAAGTGTGGAAGAGCTTCAGGAGTTTAAGAA 98492

QY 594 sProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluValPr 614

Db 98493 ACCTATGAAAAAGGAG--ATGGA AAAAGATGGAGATGATCTTTTGAAGGTGAAGTTCC 98549

QY 614 oGlnAsnAspThrValIleGlyIleThrProSerSerPheAspThr-----HisPhear 632

Db 98550 CCCAAATGATACCATGATCGGGATTATACCGAGCCACTTTGATGC-GATGCACATTTCCA 98608

QY 632 gSerProSerSerValGlySerProProValLeuTyrMet 646

Db 98609 AAGTCCITCAAGTAGTGTGGGCTCCCCACCTGTGTTACACCTG 98651

RESULT 7

AC013444 192815 bp DNA linear PRI 09-JAN-2002

DEFINITION Homo sapiens BAC clone RP11-470B22 from 2, complete sequence.

AC013444

AC013444.8 GI:18093267

VERSION HTG.

KEYWORDS Homo sapiens (human)

SOURCE ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 192815)

AUTHORS Sulston,J.E. and Waterston,R.

Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792

PUBMED 9847074

REFERENCE 2 (bases 1 to 192815)

AUTHORS Waligorski,J., Haakenson,W. and Spalding,L.

TITLE The sequence of Homo sapiens BAC clone RP11-470B22

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 192815)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (11-NOV-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 192815)

AUTHORS Waterston,R.

TITLE Direct Submission

JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT On Jan 9, 2002 this sequence version replaced gi:13518264.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0470B22  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC024616; the clone sequenced to the right is RP11-190J23. Actual start of this clone is at base position 1 of RP11-470B22; actual end is at base position 192815 of RP11-470B22.

Unresolved homopolymeric run from base position 164727 to 164753.

FEATURES

source

Location/Qualifiers

1..192815  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"

/clone="RP11-470B22"  
/clone\_lib="RPCI-11"

repeat\_region

1..575  
/rpt\_family="L1"

repeat\_region

611..878  
/rpt\_family="ERV1"

repeat\_region

879..1032  
/rpt\_family="(TAGA)n"

repeat\_region

1046..1092  
/rpt\_family="ERV1"

repeat\_region

1987..2346  
/rpt\_family="L2"

repeat\_region

2737..2972  
/rpt\_family="L1"

repeat\_region

3702..3739  
/rpt\_family="AT\_rich"

misc\_feature

4870..4891  
/note="match to EST BE071256 (NID:g8416143)"

repeat\_region

6880..6992  
/rpt\_family="MIR"

repeat\_region

8677..8934  
/rpt\_family="Alu"

repeat\_region

8773..8800  
/rpt\_family="AT\_rich"

repeat\_region

8955..8989  
/rpt\_family="AT\_rich"

repeat\_region

9207..9247  
/rpt\_family="L2"

repeat\_region

9342..9773  
/rpt\_family="L2"

repeat\_region

9977..10011  
/rpt\_family="(TTTA)n"

repeat\_region

10976..11265  
/rpt\_family="Alu"

repeat\_region

12178..12207  
/rpt\_family="(T)n"

repeat\_region

12362..12390  
/rpt\_family="AT\_rich"

repeat\_region

12702..12731  
/rpt\_family="(TG)n"

repeat\_region

12731..12758  
/rpt\_family="(GA)n"

repeat\_region

12758..12785  
/rpt\_family="(TG)n"

repeat\_region

13033..13054  
/rpt\_family="(CAAAA)n"

repeat\_region

13195..13699  
/rpt\_family="L1"

repeat\_region

13543..13567  
/rpt\_family="AT\_rich"

repeat\_region

14142..14305  
/rpt\_family="MIR"

repeat\_region

14500..14534  
/rpt\_family="AT\_rich"

repeat\_region

14617..14735  
/rpt\_family="MIR"

repeat\_region

14736..14836  
/rpt\_family="MER2\_type"

repeat\_region

14837..15106  
/rpt\_family="Alu"

repeat\_region

15107..15317  
/rpt\_family="MER2\_type"

repeat\_region

15318..15385  
/rpt\_family="MIR"

repeat\_region

15578..16066  
/rpt\_family="ERV1"

repeat\_region

16072..16112  
/rpt\_family="AT\_rich"

repeat\_region

18192..18222  
/rpt\_family="AT\_rich"

repeat\_region

18612..18899  
/rpt\_family="Alu"

misc\_feature

19181..19485  
/note="match to EST R89091 (NID:g953918) YQ02f04.s1"

misc\_feature

19879..21714  
/note="similar to Rattus norvegicus EST Y16973 (NID:g4727035)"

misc\_feature

19890..20284  
/note="match to EST R89197 (NID:g954024) YQ02f04.r1"

misc\_feature

19915..20369  
/note="similar to Homo sapiens EST BG535886 (NID:g13527431)"

misc\_feature

19923..20378  
/note="similar to Homo sapiens EST AA213789 (NID:g1812416) z91a11.s1"

misc\_feature

19923..20369  
/note="similar to Homo sapiens EST BG428305 (NID:g13334811)"

misc\_feature

20091..20378  
/note="similar to Homo sapiens EST AI742153 (NID:gs110441) wg49f01.x1"

misc\_feature

20158..20664  
/note="similar to Homo sapiens EST AV703279 (NID:g10720608)"

misc\_feature

20243..21062  
/note="similar to Mus musculus EST BG295610 (NID:g13057417)"

misc_feature	20412..21044 /note="similar to Homo sapiens EST BE734354 (NID:g10148346)"				
misc_feature	20567..20666 /note="similar to Homo sapiens EST BG778634 (NID:g14048951)"				
misc_feature	20592..20963 /note="similar to Homo sapiens EST AA481295 (NID:g2210847) zv02e05.x1"				
misc_feature	20703..21362 /note="similar to Homo sapiens EST BG502871 (NID:g13464388)"				
misc_feature	20798..21056 /note="similar to Homo sapiens EST AV649508 (NID:g9870522)"				
misc_feature	20862..21269 /note="similar to Homo sapiens EST AA191111 (NID:gl779805) zp86a03.x1"				
misc_feature	20930..21593 /note="similar to Homo sapiens EST BF203684"				
Alignment Scores:					
Pred. No.:	9.65e-183	Length:	192815		
Score:	2039.00	Matches:	453		
Percent Similarity:	77.71%	Conservative:	56		
Best Local Similarity:	69.16%	Mismatches:	120		
Query Match:	59.92%	Indels:	32		
DB:	9	Gaps:	13		
US-09-932-678-2 (1-651) x AC013444 (1-192815)					
Qy	2	AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerSerAlaVal	21		
Db	19815	GCAACCAAGGCTGCTTCACATGTTGCCGGGAA-GCAGAGGCTTCAGCTTCTGTGGTC	19873		
Qy	22	LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe	41		
Db	19874	AAGAAGCTGGGTATGTCATGACCAGTATTTACAGATATGCTTGCAATTCAGAATGATTTT	19933		
Qy	42	PheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeu	61		
Db	19934	TCCAGCTCTCCCCAAGAAAATTGTTTCATCTTGATGGAACGGGGACAGAGCTTTGCCG	19993		
Qy	62	LysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspPro	81		
Db	19994	AAGGACAAAGAGTGAAACAAATTAACTTTGAGCTGTTGAAGAGCCAGTTGTGAGATTCA	20053		
Qy	82	AspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyr	101		
Db	20054	AACATAAAGGATGACCAGATCTTGAACCTGGCTGTGAAAATTTGCTCTTCTATCATGTAC	20113		
Qy	102	LeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArg	121		
Db	20114	TTGGCAAAAGACTTTGAGCAGCTCCTCAGTATTACATTGAGATTGCCTTGGTTGAATAGA	20173		
Qy	122	SerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThr	141		
Db	20174	ATTCAAAACAGTGGTGAAGAGTATTTGGCTTCTCTGGGTACTCTTGTAACAAC-AGA---	20229		
Qy	142	ValPheLeuArgProCysLeuSerMetIleAlaSerHisPheVal---ProProArgVal	160		
Db	20230	CTCTTCTTAGCCCAAGTCTCAGCATAATTGCTTCCCAATTTGGCGTACCCTCCTTGAATG	20289		
Qy	161	IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAsnLeuPro	180		
Db	20290	ATCATTAAGAAGAGCCATGTAATGTTTCAGATTCTGATGATGGAGATGATTATCTTCCT	20349		
Qy	181	AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr	200		
Db	20350	GCAGATTTTGACACACGTCATACAACTTTACAAATAACAGCAAGATATGTCCCACTGGCA	20409		
Qy	201	ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg	220		

Db	20410	CTGTGGTTTCTTTTGGCCCATACTGGTGGAAACAATTTCCATTTGTT---CAATCAGAGAGA	20466
Qy	221	ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg	240
Db	20467	ACACTGGAATGTTACATTCATAACTTACTAAAGATTAGTATATACTTTCCAACCTTGAGG	20526
Qy	241	HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg	260
Db	20527	AATGAAATCTGAAGCTTATTATAAAATCTATTCAAATTTGGATGGATACATTTCTGG	20586
Qy	261	GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr	280
Db	20587	CATGATATTGAAGATGCTGAGGAAACACAGAGCTCAAACTTGTGGTGGACAGATTTCCATG	20646
Qy	281	GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro	300
Db	20647	GAATGATTTGTTCAATATGAACGAAGATGAAGAGACT-----GTTTCTAAGGCTGACTCC	20700
Qy	301	GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu	320
Db	20701	AAATGGCTCGGTACATGGTGCACCCCTGTAGCTGAATGCCTGGACACCTGCTGTCTTTG	20760
Qy	321	ValLeuSerTyrMetLysAspValCysTyrValAsp-----GlyLysValAspAsnGly	338
Db	20761	CTTTTGCAGTACTTGAAGTATGCTGCTATGTAAA-ATACATGGTAAGTTGATAACAAC	20819
Qy	339	LysThrLysAspLeu-----TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeu	356
Db	20820	AAAATAAGGATTT-ATAATATATCACAAACTGATAAACATATTTGACAAA-----	20869
Qy	357	ProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuLeuCysSerPheLysLeu	376
Db	20870	---CCCCT-GCCTCCTGTCACGTACAGTTTTTTATGTTTTACCTCTGTGGCTTAAAAATTG	20925
Qy	377	GlyPheAlaGluAlaPheLeuGluHisLeuTrp---LysLysLeuGlnAspProSerAsn	395
Db	20926	CGACTCTCAGAAGCATTTTTGGAAACGTCCTCTGGGAAAAAAAATGCAGGATCCAAATAAT	20985
Qy	396	ProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLys	415
Db	20986	CCTGTCATCATCAGGTGAGCTGCTGGAATATATATTGGAAGCTCTTTGGCAAAAGCTAAA	21045
Qy	416	PheIleProLeuIle---ThrValLysSerCysLeuAspLeuValAsnTrpLeuHis	434
Db	21046	TTCATTCCTCTCATTAACAATAATAAAATCATGCCTAGGTCTTTTCGGTTAAGTGGCCACAC	21105
Qy	435	IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly	454
Db	21106	AAATACCTTACTAACCCAGGATCTGAGAAATGAAGGCTTTTGTGATATTGCTCTCCATCAA	21165
Qy	455	ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu	474
Db	21166	ACATTTTACTCAGCCTTCCAAGCTGTGTCCACACCTTTGTTTTTAGACACAAGCAACTT	21225
Qy	475	LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle	494
Db	21226	TTAAGCGTAAACCTGAAAGAAGGTTTGAGATGTCTTCAGTCCTACATTTTAAAGGATA	21285
Qy	495	ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla	514
Db	21286	GTGACTAGCAGCTAAATCCTCTGAATATTTGCCCTGCCCTCAGTGGTTAACTTATTTGCT	21345
Qy	515	AlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn-Ar	534
Db	21346	GCTGTCACGAATAAATATCAGCTAGTC---TGCTACACTATCATTTGAGAGAAACAACGGC	21402
Qy	534	gGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAs	554
Db	21403	C-----AGAAATACAGCTGGAGGAGACTCGGTGTGAACCTGCACCAA	21444
Qy	554	nProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPheIl	574
Db	21445	CCCATTGGACGCCCTTCTTCCCCTTTGATCTCTTGTGCGCTTAAGCAGTCAAGGAAAAGTCAT	21504





```
QY      490 nPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValVa 510
Db      1304 TTTTGAGCGGATAGTGTGATGAGCCAGCTAAATCCCTGAAGATTGGCTGCCCTCAGTGGT 1363

QY      510 lAsnPhePheAlaAlaIleThr 517
Db      1364 TAACITTTTGTGCTGCAATCACa 1385

RESULT 9
AX274907
LOCUS      AX274907      1461 bp      DNA      linear      PAT 29-OCT-2001
DEFINITION Sequence 172 from Patent WO0172777.
ACCESSION  AX274907
VERSION     AX274907.1  GI:16547539
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Hillman,J.L., Baughn,M.R., Yue,H., Lal,P., Lu,D.A., Patterson,C.,
            Azimzai,Y., Bandman,O., Tang,Y.T., Mathur,P., Shah,P., Au-Young,J.
            and Reddy,R.
TITLE       Transcription factors
JOURNAL     Patent: WO 0172777-A 172 04-OCT-2001;
            Incyte Genomics, Inc. (US)
FEATURES
            source
            1..1461
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="Incyte ID No: 2646274CB1"

ORIGIN
Alignment Scores:
Pred. No.:      6.51e-98      Length:      1461
Score:          1124.00      Matches:      256
Percent Similarity: 55.44%      Conservative: 14
Best Local Similarity: 52.57%      Mismatches:  52
Query Match:    33.03%      Indels:      166
DB:             6      Gaps:      10

US-09-932-678-2 (1-651) x AX274907 (1-1461)

QY      195 ArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPhe 214
Db      263 AAATACAAAATTAGGGCACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCATTT 322

QY      215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerVal 234
Db      323 GTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCAATACTTACTAAGGATTAGTGA 382

QY      235 TyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeu 254
Db      383 TATTTTCCAACTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAACTACTCAAGCTG 442

QY      255 AspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCys 274
Db      443 GATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAAAATCAAACCTTGT 502

QY      275 GlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHis 294
Db      503 GGTGGGACAGATCCACGGAAGGATTGTTTAAATAG-----

QY      295 GluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeu 314
Db      538 -----

QY      315 AspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLys 334
Db      538 -----

QY      335 ValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeu 354
Db      538 -----

QY      355 LeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPhe 374
Db      538 -----

QY      375 LysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSer 394
Db      539 -----GGATTGCGCAGAGGCATTTTGGAAACATCTTTGGAAAAACCTTCAGGATCCAAGT 592

QY      395 AsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAla 414
Db      593 AATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTGGCAAGAGCT 652

QY      415 LysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHis 434
Db      653 AAATTTATTCTCTTATTACTGTATAAACCATGCCTAGATCTTTTGGTTAACTGGCTGCAC 712

QY      435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454
Db      713 ATATACCTTAATAACCAAGGATTCGGGAACAAAGGCATCTCGGATGTTGCTCTCCATGGA 772

QY      455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
Db      773 CCATTTTACTCAGCCTGCCAAGCTGTGTCTTACACCTTGTGTTTTTAGACACAAGCAGCTT 832

QY      475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494
Db      833 TTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCCTCAGAGTCTGAATTTTGAGCGGATA 892

QY      495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla 514
Db      893 GTGATGAGCCAGCTAAATCCCTGAAGATTGCCTGCCCTCAGTGGTTAACTTTTGTGCT 952

QY      515 AlaIleThrAsnLysTyr-GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnAr 534
Db      953 GCAATCACA---AAGATGAAGACTTGIG-----

QY      534 gGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAs 554
Db      978 -----GATATGGATGGTGGTGGT-----TGCACAAC 1006

QY      554 nProLeuAspThrPhePheProPheAspProCysValLeuLys----- 568
Db      1007 AAT-ATCAATTTATTTATACCACCTGAACCGTGCATTTCAAAATGGTTAAGATGTTAAGT 1065

QY      569 -----ArgSerLysLysPheIleAspPr 576
Db      1066 TTTATGTTGTGTATTTTACCATAATAAAAAAATTGTAGAGGAAAAAACAGTCTGCCTCC 1125

QY      576 oIleTyrGlnValTrpGlu-----AspMetSerAlaGluGl 588
Db      1126 ACTTTTG---ATATGGAGCTGCTAACATCTCCACCTTCGCTCTCCCTTCTGCCCCACA 1182

QY      588 uLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAspAspPh 608
Db      1183 TCTGGCAAGCTAAGAAAGCCT----- 1204

QY      608 eLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer----- 625
Db      1205 -----GCTGCTCTCTCCTCTGGCACAGCTGGAAATTCATACCCACAAGCCCTA 1254

QY      626 ----SerPheAspThrHisPheArgSerProSerSerSerValGlySerProProValLe 644
Db      1255 GCCTTCCCAACAGACCCACATTTCATCCCATCCCATCGCATCCCCATCCC----- 1306

QY      644 uTyrMetGlnProSerPro 650
Db      1307 -----CATCCCCATCCC 1318

RESULT 10
```



BC006441	2627 bp	mrna	linear	PRI 02-JUL-2003
LOCUS	Homo sapiens similar to RNA polymerase I transcription factor RRN3, mRNA (cDNA clone IMAGE:3351791), partial cds.			
DEFINITION	Homo sapiens similar to RNA polymerase I transcription factor RRN3, mRNA (cDNA clone IMAGE:3351791), partial cds.			
ACCESSION	BC006441			
VERSION	BC006441.1	GI:13623642		
KEYWORDS	.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 2627) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.			
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
MEDLINE	22388257			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 2627)			
AUTHORS	Strausberg,R.			
TITLE	Direct Submission			
JOURNAL	Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>			
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabs-r@mail.nih.gov">cgabs-r@mail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a> Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.			
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAL Plate: 18 Row: d Column: 24. Location/Qualifiers 1. .2627 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:3351791" /tissue_type="Eye, retinoblastoma" /clone_lib="NIH_MGC_16" /lab_host="DH10B-R" /note="Vector: pOTB7"			
source				

ORIGIN			
Alignment Scores:			
Pred. No.:	6.11e-86	Length:	2627
Score:	1003.00	Matches:	233
Percent Similarity:	53.59%	Conservative:	13
Best Local Similarity:	50.76%	Mismatches:	48
Query Match:	29.47%	Indels:	166
DB:	9	Gaps:	10
US-09-932-678-2 (1-651) x BC006441 (1-2627)			
QY	223	GluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGlu	242
Db	1509	GAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTTTCCAACCTTGAGGCATGAA	1568
QY	243	IleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGly	262
Db	1569	ATTCTGGAGCTTATTATTGAAAAACTACTCAAGCTGGATGTGAATGCATCCCGCAGGGT	1628
QY	263	IleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly	282
Db	1629	ATTGAAGATGCTGAAGAAACAGCAAAATCAAACTTGTGTGGACAGATTCCACGGAAGGA	1688
QY	283	LeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArg	302
Db	1689	TTGTTTAAATATG-----	1700
QY	303	LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu	322
Db	1700	-----	1700
QY	323	SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp	342
Db	1700	-----	1700
QY	343	LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCys	362
Db	1700	-----	1700
QY	363	HisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaphe	382
Db	1701	-----GGATTGCGAGAGGCATTT	1718
QY	383	LeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleAargGlnAla	402
Db	1719	TTGGAACATCTTTTGAAAAACTTGCAGGATCCAAGTAATCTGCCATCATCAGGCAGGCT	1778
QY	403	AlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal	422
Db	1779	GCTGGAATTATATTGGAAAGCTTTTGGCAAGAGCTAAATTTATTTCTTATTACTGTA	1838
QY	423	LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSer	442
Db	1839	AAACCATGCGTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAAGGATTCTG	1898
QY	443	GlyThrLysAlapheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla	462
Db	1899	GGAACAAAGGCATTTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCT	1958
QY	463	ValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGly	482
Db	1959	GTGTTCTACACCTTTGTGTTTGTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT	2018
QY	483	LeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeu	502
Db	2019	TTGCAGTATCCTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTG	2078
QY	503	LysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr-GlnLe	522
Db	2079	AAGATTGCCTGCCCTCAGTGGTTAACTTTTGTGCTGCAATCACA---AAGATGAAGACT	2135
QY	522	uValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSe	542

Db	2136	TGTG-----	-----GA	2141
QY	542	rThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPh	562	
Db	2142	TATGGATGGTGTATGATGTT-----TGCACAACAAT-ATCAATTATTTATACCAC	2191	
QY	562	eAspProCysValLeuLys-----	568	
Db	2192	TGAACCGTGCACCTCAAAATGGTTAAGATGGTAAGTTTATGTTGTGTATTTTACCATAA	2251	
QY	569	-----ArgSerLysLysPheIleAspProIleTyrGlnValTrpGlu-----	582	
Db	2252	TAAAAAAATTTGTAGAGGAAAAACAGTCTGCCCTCCACTTTG--ATATGGGACTGCTA	2308	
QY	583	-----AspMetSerAlaGluGluLeuGlnGluPheLysLysProMe	596	
Db	2309	ACATCTTCCACCTTCCCTCTCCCCCTCTGCCCCACATCTGGGCAAGCTAAGAAAGCCT--	2366	
QY	596	tLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluValProGlnAs	616	
Db	2367	-----	-----GCTGCTCTCTCCTC	2380
QY	616	nAspThrValIleGlyIleThrProSer-----SerPheAspThrHisPheAr	632	
Db	2381	TGGCACCAGCTGGAAATTCATACCCCAACAGCCCTAGCCCTCCACCAGACCCACATTTC	2440	
QY	632	gSerProSerSerValGlySerProProValLeuTyrMetGlnProSerPro	650	
Db	2441	ATCCCCATCCCCATCGCATCCCCCATCCC-----	-----CATCCCCCATCCC	2480
RESULT 11				
AX713860				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
ORIGIN				
Alignment Scores:				
Pred. No.:				
Score:				
Percent Similarity:				
Best Local Similarity:				
Query Match:				
DB:				
US-09-932-678-2 (1-651) x AX713860 (1-2493)				
QY	374	PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro	393	
Db	703	TATCGACGGGATTCGACAGAGCATTTTGGAAACATCTTTGGAAAAAATTCAGGATCCA	762	
QY	394	SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg	413	

Db	763	AGTAATCTGCTCCATCATCAGGCAGGCTGCTGGAATATATATATGGAAGCTTTTGGCAAGA	822	
QY	414	AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuValAsnTrpLeu	433	
Db	823	GCTAAATTTATTCCTCTTATTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTG	882	
QY	434	HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis	453	
Db	883	CACATATACCTTAATAACAGGATTCGGGAACAAAGGCATTTCTGCAATGTTGCTCTCCAT	942	
QY	454	GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln	473	
Db	943	GGACCAATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAG	1002	
QY	474	LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg	493	
Db	1003	CTTTTGAGCGGAAACCTGAAAGAAGGTTTGAGTATCTCAGAGTCTGAATTTTGAGCGG	1062	
QY	494	IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe	513	
Db	1063	ATAGTATGATGAGCCAGCTAAATCCCCTGAAGATTTGCTGCTGCCCTCAGTGGTTAACTTTT	1122	
QY	514	AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn	533	
Db	1123	GCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGTGTACCATCATCAGAGGAACAAT	1182	
QY	534	ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr	553	
Db	1183	CGCCAGATGCTGCCAGTCATTAGGAGTACCACCTGGAGGAGACTCAGTGCAGACCTGCACA	1242	
QY	554	AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArg	569	
Db	1243	AACCCACTGGACACCTTCGTCCCTTTGATCCCTGTGTGCTGAAGAGG	1290	
RESULT 12				
AK055742				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				

Location/Qualifiers





/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6403754"  
/tissue type="Brain, mouse 12.5 dpc"  
/clone\_lib="NIH BMAP\_FD0"  
/lab\_host="DH10B"  
/note="Vector: pYX-ASC"  
<1..2243  
/gene="AL023001"  
/note="synonym: Rrn3"  
/db\_xref="LocusID:106298"  
/db\_xref="MGI:2146525"  
<1..647  
/gene="AL023001"  
/codon\_start=3  
/product="AL023001 protein"  
/protein\_id="AAH55781.1"  
/db\_xref="GI:33585914"  
/db\_xref="LocusID:106298"  
/db\_xref="MGI:2146525"  
/translation="KAFCDVALHGPFYSACQAVFTVVRHKQLLSGNLKQGLQYLQS  
LNFERIVLSQLNPLKICLPQVNVFFAATNKYQLVFCYTIMERNSRQMLPVIRSTAGG  
DSVQCTNPLDTFFPDPCLVLRSKKPIDPIYQIWEDESAEELQEFKKSTKKQVVEDE  
DDDFLKGEVPQSDTWTGLTPSSFDTHQSPSSSVGPPVLYIPGQSPLLTRIYD"  
27..434  
/gene="AL023001"  
/note="Region: RNA polymerase I specific transcription  
initiation factor RRN3. This family consists of several  
eukaryotic proteins which are homologous to the yeast RRN3  
protein. RRN3 is one of the RRN genes specifically  
required for the transcription of rDNA by RNA polymerase I  
(Pol I) in Saccharomyces cerevisiae"  
/db\_xref="CDD:pfam05327"

ORIGIN

Alignment Scores:  
Pred. No.: 6,21e-85 Length: 2243  
Score: 991.50 Matches: 187  
Percent Similarity: 94.71% Conservative: 10  
Best Local Similarity: 89.90% Mismatches: 10  
Query Match: 29.14% Indels: 1  
DB: 10 Gaps: 1

US-09-932-678-2 (1-651) x BC0555781 (1-2243)

QY 445 LysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPhe 464  
Db 3 AAGGCTTTTGTGACGTTGCACCTCCATGGACCATTTTATTTCAGCCTGCCAAGCTGTGTTTC 62

QY 465 TyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGln 484  
Db 63 TACACTGTGTTTTTAGACACAAGCAGCTTTTGAGTGGAAACTTGAAGCAAGGTCTACAG 122

QY 485 TyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIle 504  
Db 123 TATCTTCAGAGTCTAAATTTTGAGCGCATTTGTGTGAGCCAGCTGAACCCACTGAAGATC 182

QY 505 CysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyrGlnLeuValPhe 524  
Db 183 TGCCTGCCGCAAGTGGTTAATTTCTTTGCTGCTATCACAAATAAATACCAGCTGGTGTTTC 242

QY 525 CysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAla 544  
Db 243 TGCTACACCATCATGGAGAGGAACAGTCGCCAGATGCTCCCGTTATCCGCAGCACCGCT 302

QY 545 GlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPheProPheAspPro 564  
Db 303 GGTGGAGACTCCGTGCGACACCTGCACCAACCCACTGGACACTTCTTCCCTTTGACCCCT 362

QY 565 CysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspMet 584  
Db 363 TGTGTGCTTAAGAGGTCAAAGAAGTTTCATTGATCCCTATTATCAGATTGGGAAGATGGG 422

QY 585 SerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGlu 604  
Db 423 AGTGTGAAGAGCTTCAGAGATTTAAGAAATCTACTAAAAAGCAGGTAGTGGAGGATGAA 482

QY 605 AspAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrPro 624  
Db 483 GATGATGACTTTTGAAGGCGAGGTGCCAGAGTGACACAGTGGCCTTACTCCG 542

QY 625 SerSerPheAspThrHisPheArgSerProSerSerValGlySerProProValLeu 644  
Db 543 AGCTCCTTTGATACCCACTTCCAAAGTCTTCCAGTAGTGTGGGCTCCCTCCTGTCTG 602

QY 645 TyrMetGln---ProSerProLeu 651  
Db 603 TATATACGAGGCCAGTCTCCACTC 626

RESULT 15  
AK126166  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,16e-84 Length: 3645  
Score: 986.00 Matches: 212  
Percent Similarity: 62.54% Conservative: 5  
Best Local Similarity: 61.10% Mismatches: 17  
Query Match: 28.97% Indels: 114  
DB: 9 Gaps: 4

US-09-932-678-2 (1-651) x AK126166 (1-3645)

Search completed: July 18, 2004, 10:56:28  
Job time : 7164 secs

```
QY      223  GluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGlu 242
Db      2588  GAATGTTACGTTTCATAACTTACTAAGGATTAGTGATATATTTCCAACTTGAGGCATGAA 2647

QY      243  IleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGly 262
Db      2648  ATTCTGGAGCTTATTATTGAAAACTACTCAAGCTGATGTGAATGCATCCCGCAGGGT 2707

QY      263  IleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly 282
Db      2708  ATTGAAGATGCTGAAGAAACAGCAAAATCAAACCTTGTTGGGACAGATTCCACGGAAGGA 2767

QY      283  LeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArg 302
Db      2768  TTGTTTAATATG----- 2779

QY      303  LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322
Db      2779  ----- 2779

QY      323  SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp 342
Db      2779  ----- 2779

QY      343  LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCys 362
Db      2779  ----- 2779

QY      363  HisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPhe 382
Db      2780  -----GGATTTCGACAGGCATTT 2797

QY      383  LeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAla 402
Db      2798  TTGGAACATCTTTGGAAAACTTGCAGGATCCAAGTAATCTGCCATCATCAGGCAGGCT 2857

QY      403  AlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal 422
Db      2858  GCTGGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTATTCTCTTATTACTGTA 2917

QY      423  LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSer 442
Db      2918  AAACCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCTG 2977

QY      443  GlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462
Db      2978  GGAACAAAGGCATTCTGCGATGTGCTCTCCATGGACCACTTTTACTCAGCCTGCCAAGCT 3037

QY      463  ValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
Db      3038  GTGTTCTACACCTTTGTTTTTAGACACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 3097

QY      483  LeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeu 502
Db      3098  TTGCAGTATCCTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCTTG 3157

QY      503  LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr-GlnLe 522
Db      3158  AAGATTTGCCCTGCCCTCAGTGGTTAACTTTTGTGTCATCACA---AAGATGAAGACT 3214

QY      522  uValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSe 542
Db      3215  TGTG-----GA 3220

QY      542  rThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPh 562
Db      3221  TATGGATGGTGGTGATGGT-----TGACACAACAAAT-ATCAATTATTATTATACCAC 3270

QY      562  eAspProCysValLeuLys 568
Db      3271  TGRACCGTGCACCTTCAAAA 3289
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 08:21:17 ; Search time 4405 Seconds  
(without alignments)  
4413.230 Million cell updates/sec

Title: US-09-932-678-2  
Perfect score: 3403  
Sequence: 1 MAAPLLHTRLPGDAAASSA.....RSPSSVGSPPVLYMQPSPL 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spo01/US0932678/runat\_15072004\_103604\_5218/app\_query.fasta\_1.839  
-DB=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0932678@cgn\_1\_1\_3437\_@runat\_15072004\_103604\_5218 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2966.5	87.2	3457	11	AK080397	AK080397 Mus muscu
2	2742.5	80.6	2136	11	AK080680	AK080680 Mus muscu
3	1755.5	51.6	1800	14	Y16973	Y16973 RNY16973 Ra
4	1290	37.9	896	13	BQ214727	BQ214727 AGENCOURT
5	1242.5	36.5	894	14	CD359064	CD359064 AGENCOURT
6	1223	35.9	889	13	BU510623	BU510623 AGENCOURT
7	1208	35.5	828	13	BU610641	BU610641 UI-M-FCO-
8	1207.5	35.5	788	12	BG502871	BG502871 602550430
9	1125.5	33.1	996	10	BF203684	BF203684 601866331
10	1120	32.9	754	14	CA511975	CA511975 UI-R-FUJ-
11	1116.5	32.8	769	14	CF738084	CF738084 UI-M-HD0-
12	1111	32.6	682	14	CB150625	CB150625 K-EST0207
13	1105	32.5	733	12	BI561523	BI561523 603256222
14	1100.5	32.3	1037	12	BG428305	BG428305 602498983
15	1094	32.1	757	10	BE734354	BE734354 601565596
16	1048	30.8	651	14	CF908118	CF908118 A0519A10-
17	1040	30.6	671	9	AV703279	AV703279 AV703279
18	1040	30.6	1122	13	BQ066201	BQ066201 AGENCOURT
19	1008.5	29.6	827	13	BU129543	BU129543 603118483
20	1003.5	29.5	2659	11	BC037959	BC037959 Homo sapi
21	1000.5	29.4	828	12	BI761160	BI761160 603043657
22	989.5	29.1	736	12	BI464721	BI464721 603202341
23	975	28.7	580	12	BM564978	BM564978 iJ71d11.Y
24	968.5	28.5	878	13	BX743535	BX743535 BX743535
25	968	28.4	609	14	CF902795	CF902795 A0346A07-
26	966	28.4	563	10	AW408066	AW408066 UI-HF-BM0
27	966	28.4	727	12	BI549394	BI549394 603190049
28	964	28.3	551	12	BM786521	BM786521 K-EST0065
29	961.5	28.3	690	9	AJ446028	AJ446028 AJ446028
30	960	28.2	712	13	BY733332	BY733332 BY733332
31	955	28.1	568	9	AL708660	AL708660 DKFZp886K
32	951	27.9	645	13	BQ638571	BQ638571 hd25a12.Y
33	946	27.8	654	14	CF737683	CF737683 UI-M-HD0-
34	941	27.7	571	14	CA563382	CA563382 K0313C08-
35	937.5	27.5	750	9	AJ446022	AJ446022 AJ446022
36	936	27.5	612	13	BQ771414	BQ771414 UI-M-FIO-
37	935	27.5	990	12	BG167061	BG167061 602344740
38	928.5	27.3	1045	12	BG297178	BG297178 602395174
39	923.5	27.1	670	13	BY721737	BY721737 BY721737
40	921	27.1	753	14	CB266771	CB266771 1005677 H
41	919.5	27.0	916	13	BX455091	BX455091 BX455091
42	911.5	26.8	703	13	BU703050	BU703050 UI-M-FD0-
43	909	26.7	598	14	CF910307	CF910307 A0547H07-
44	902.5	26.5	842	9	AJ446021	AJ446021 AJ446021
45	902	26.5	564	10	AW958173	AW958173 EST370243

ALIGNMENTS

RESULT 1  
AK080397  
LOCUS  
DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630092F02 product:RRN3 homolog [Homo sapiens], full insert sequence.  
ACCESSION AK080397  
VERSION AK080397.1 GI:26099204  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AK080397 3457 bp mRNA linear HTC 19-SEP-2003  
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630092F02 product:RRN3 homolog [Homo sapiens], full insert sequence.

**REFERENCE**

**AUTHORS**  
Carninci, P., and Hayashizaki, Y.  
**TITLE**  
High-efficiency full-length cDNA cloning  
**JOURNAL**  
Meth. Enzymol. 303, 19-44 (1999)  
**MEDLINE**  
99279253  
**PUBMED**  
10349636

**REFERENCE**

**AUTHORS**  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**TITLE**  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
**JOURNAL**  
Genome Res. 10 (10), 1617-1630 (2000)  
**MEDLINE**  
20499374  
**PUBMED**  
11042159

**REFERENCE**

**AUTHORS**  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
**TITLE**  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
**JOURNAL**  
Genome Res. 10 (11), 1757-1771 (2000)  
**MEDLINE**  
20530913  
**PUBMED**  
11076861

**REFERENCE**

**AUTHORS**  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
**TITLE**  
Functional annotation of a full-length mouse cDNA collection  
**Nature** 409, 695-690 (2001)

**REFERENCE**

**AUTHORS**  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
**TITLE**  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**Nature** 420, 563-573 (2002)

**REFERENCE**

**AUTHORS**  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
**TITLE**  
Direct Submission  
**JOURNAL**  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT**  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.  
please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers  
1. .3457

**FEATURES**  
source

---

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FANTOM.DB:A630092F02"
/db_xref="MGI:2407489"
/db_xref="taxon:10090"
/clone="A630092F02"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
/misc_feature misc_feature
14. .1987
/note="RRN3 homolog [Homo sapiens] (SPTR|Q9NYV6, evidence: FASTY, 87.8%ID, 92%length, match=1791) putative"

ORIGIN
Alignment Scores:
Pred. No.:      0          Length:    3457
Score:         2966.50     Matches:   570
Percent Similarity: 93.13%   Conservative: 40
Best Local Similarity: 87.02% Mismatches: 39
Query Match:    87.17%     Indels:    6
DB:             11        Gaps:       2

US-09-932-678-2 (1-651) x AK080397 (1-3457)

QY      1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerSerAla 20
Db      14 ATGGCGCCCCGCTGCTTCACACGCCGCTTGTCGGGCGATGTGACGGCAGCGCCTCTGCC 73

QY      21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40
Db      74 ACG-----CTGAGTGCCTCGCGACTGGGCTTTCCGATATGCTCGCGTAGAGCGCAT 127

QY      41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60
Db      128 TTCTTCAATTCTCTCCGAAAAAACTGTTTCGGTTTGGCGGAACCTGTGACAGAAGTCTTA 187

QY      61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLysAsnGlnLeuLeuAsp 80
Db      188 CTGAATAACA AAAAGGGTGAAA CAATGACTTAGAGTTGTTGAAGAAC CCAGCTGTCTGAT 247

QY      81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
Db      248 CCTGATATAAAGGATGACCAGATCATTA ACTGGCTACTAGAAATTCGTTCTCTGTCATG 307

QY      101 TyrLeuThrLysAspPheGluGlnLeuLeuSerIleIleLeuArgLeuProTrpLeuAsn 120
Db      308 TACTTGACAAAAGACTTTTGAGCAACTTATAAACATCATATATGAGATTGCCCTGGTTGAAT 367

QY      121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140
Db      368 AGAAGTCAGAGAGTGGTGAGGAGTATT TGGCTTTTCTGGTAAATCTTGIGTGTGCACAG 427

QY      141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160
Db      428 ACTGCTTCCTTAGACCATGTCTCAGCATGATTGCGTCTCATTTTGTACCTCCCAGGATA 487

QY      161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180
Db      488 ATTGTCAGGAAGGTGGCATAGATGTTTCAGATTCTGATGACGAAGATGACAATCTTCCT 547

QY      181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
Db      548 GCAATTTTGGACACATGTACAGAGCCCTTGCAAATAATAACAAGATATGTCCCATCGACA 607

QY      201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220
Db      608 CCATGGTTTCTAATGCCAATACTGGTGGAGAAATTCCTGGTGTGAGGAAGTCCGAGAGA 667

QY      221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240
Db      668 ACATTGGAATGTTAMGTTCAATACTTATTAGGATAAGTTTATATATTTCCCAACTTGAGG 727

```



Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gscl.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/.

FEATURES

SOURCE

1. .2136  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:A830089A07"  
/db\_xref="MGI:2409380"  
/db\_xref="taxon:10090"  
/clone="A830089A07"  
/tissue\_type="cortex"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"  
40. .1905  
/note="RRN3 homolog [Homo sapiens] (SPTR|Q9NYV6, evidence:  
FASTY, 87.8%ID, 92%length, match=1791)  
putative"  
2110. .2115  
/note="putative"  
2136  
/note="putative"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2136  
Score: 2742.50 Matches: 549  
Percent Similarity: 86.51% Conservative: 41  
Best Local Similarity: 80.50% Mismatches: 56  
Query Match: 80.59% Indels: 38  
DB: 11 Gaps: 7

US-09-932-678-2 (1-651) x AK080680 (1-2136)

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerSerAla 20  
Db 40 ATGGCGGCCCGCTGCTTCACACGCGCTGTTCGGCGGATGTGACGGCAGCGCCTCTGCC 99  
QY 21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
Db 100 ACG-----CTGAGTGCCTGCGGACTGGGCTTTCGGATATGCTCGCGTTAGAGCGGAT 153  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyThrValThrGluValLeu 60  
Db 154 TTCTTCAATTCTCCTCCGAAAAAACTGTTCGGTTTGGCGAACTGTGACAGAGTCTTA 213  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
Db 214 CTGAATACAAAAAGGTGAACAAATGACTAGATTGTTGAAGAACCCAGCTGCTGAT 273  
QY 81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
Db 274 COTGATATAAAGGATGACCATCATTAACCTGGCTACTAGAAATTCGTTCCCTCTGTCTATG 333  
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn 120

Db 334 TACTTGACAAAAGACTTTGAGCAACTTATAAACATCATATTGAGATTGCCCTGGTTGAAT 393  
QY 121 ArgSerGlnThrValValGluGluTyrIleuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
Db 394 AGAAGTCAGAGAGTGGTGGAGGAGTATTTGGCTTTTCTTGGTAAATCTTGTGTGCACAG 453  
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
Db 454 ACTGTCTTCTTAGACCATGTCTCAGCATGATTCGCTCTCATTTTGTACCTCCCGAGTA 513  
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180  
Db 514 ATTGTCAAGGAAGGTGGCATAGATGTTTCAGATTCTGATGACGAAGATGACAAATCTTCT 573  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
Db 574 GCAATTTTGGACACATGTTCACAGAGCCTTGCAATAATAACAAGATATGTCCCATCGACA 633  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
Db 634 CCATGGTTTCTAATGCCAATACTGTTGGAGAAATTCCTGTTGTGAGGAAGTCCGAGAGA 693  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
Db 694 ACATTGGAATGTTATGTTTCATAACTTATTAAGGATAAGTTTATATTCCCAACTTTGAGG 753  
QY 241 HisGluIleLeuLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
Db 754 CGTGAATTTCTGGAGCTTGTATTGAAAAGCTACTCAAGTTAGATGTGAGTGTATCGCGG 813  
QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
Db 814 CAGGATATTGAAGATGCTGAAGAGAAAGCAGCTCAGACTTGTGGTGGACAGATACCACA 873  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
Db 874 GAAGGACTGTTTATATAGGATGAAGATGAGGACACTGACCTTGAGAAGAAAGCTGACCCAG 933  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
Db 934 GAGGAGCCTTAACAGATGGCTCATCCACCGCAGAGCGCCTGGACGCTCTGTGCTTG 993  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
Db 994 CTGCTGTCTCTACATTGAGGATGTCGCGGTGTGCAGGTAAATGATAACAATAAAACA 1053  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
Db 1054 AAGGATTTATACCGTGATCTGATATCCATCTTTGACAAACTTCTGTTGCCCCACATGCC 1113  
QY 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
Db 1114 TCCTGCCAATGTACAGTCTTTCATGTTTTCCTCTGACGTTCAAGTTGGGATTTCAGAAA 1173  
QY 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400  
Db 1174 GCATTTTGGAAACATCTCTGGAAAAAGTTGCAGGATCCAAATAACCCCGCCATCATCAGG 1233  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
Db 1234 CAGGCTGCTGCAAAATTATATTGGAGCTTTTGGCCAGAGCTAAATTTATTCTCTTATC 1293  
QY 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
Db 1294 ACTGTGAAGACATGCTCGGATCTCTTGGTTAACTGGCTGCACATGTACCTTACTAACCCAG 1353  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyLysProPheTyrSerAlaCys 460  
Db 1354 GATTCGGGAACAAAGGCTTTTGTGACGTTGCACT-CATGGACCACTTTTATTTCAGCCTGC 1412  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
Db 1413 CAAGCTGTGTTCTACACTGTGTTGTTTTTAGACACAAGCAGCTTTGAGTGGAAACTTGAAG 1472

QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
Db 1473 CAAGGTCTACAGTATCTTCAGAGTCTAAATTTTGAGCGCATTGTGTGAGCCAGCTGAAC 1532  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520  
Db 1533 CCACCTGAAGATCTGCCTGCCGCAAGTGTTAATTTCTTGTCTATCACAATAAATAC 1592  
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValile 540  
Db 1593 CAGCTGGTGTCTCTACACCATCATGGAGAGGAAGTCGCCAGATGCTCCCGTTATC 1652  
QY 541 ArgSer-ThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePh 560  
Db 1653 CGCAGCCACCGCTGGTGAGACTCCGTGCAGACCTGCACCAACCCACTGGACACTTCTT 1712  
QY 560 eProPheAspProCysValLeuLysArgSerLysPheIleAspProIleTyrGlnVa 580  
Db 1713 CCCCTTTGACCCCTGTGTCTTAAAGAGGTCAAAGAAGTTCAATGATCCTATTATCAGAT 1772  
QY 580 lTrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLys---- 598  
Db 1773 TTGGGAAGATGGGAGTGCTGAAGAGCTTCAGGAGTTTAAGAAATCTACTAAAAGGAATC 1832  
QY 599 -----AspIleValGluAspGluAsp-- 605  
Db 1833 TCCTGCTGGAAGAGGCAATGGCACCATTCTGAAGGGGAACCTGTCTCCAGAGAGATCT 1892  
QY 606 -AspAspPheLeuLysGlyGluVal---ProGlnAsnAspThrValIle----- 620  
Db 1893 CAATGAGAC-CTAAAAAAACTTGTGACAGCCACAGAATGAGACTAAAGCCTAAATAATGAT 1951  
QY 621 -----GlyIleThrProSer-SerPheAspThrHisP 631  
Db 1952 GGGTAGGATCCAAACACGACGAGAAATAAGGCGCTCAAAACCGGATTATCTGGCAACCGGAT 2011  
QY 631 heArgSerProSerSerValGlySerProProValLeuTyrMetGlnProSerProL 651  
Db 2012 TTCTCTCTCCCTCTCTCTCTCTC---TCTCCC-----CTCTCCCTCTCTCCCTCTCCCC 2062  
QY 651 eu 651  
Db 2063 TC 2064  
RESULT 3  
YI6973  
LOCUS YI6973 1800 bp mRNA linear EST 30-APR-1999  
DEFINITION RNYI6973 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone  
IRL573, mRNA sequence.  
ACCESSION YI6973  
VERSION YI6973.2 GI:4727035  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
REFERENCE 1 (bases 1 to 1800)  
AUTHORS Olivier,E., Soury,E., Risler,J.L., Smih,F., Schneider,K.,  
Lochner,K., Jouzeau,J.Y., Fey,G. and Salier,J.P.  
TITLE A novel set of hepatic mRNAs preferentially expressed during an  
acute inflammation in rat represents mostly intracellular proteins  
JOURNAL Genomics 57 (3), 352-364 (1999)  
MEDLINE 99263497  
PUBMED 10329001  
COMMENT On Apr 2, 1998 this sequence version replaced gi:3006069.  
Contact: E. Olivier  
U78 INSERM  
543 chemin de la Breteque, 76233 Bois-Guillaume, France  
POLYA=No.  
FEATURES Location/Qualifiers  
source 1..1800

ORIGIN  
Alignment Scores:  
Pred. No.: 6.44e-200 Length: 1800  
Score: 1755.50 Matches: 451  
Percent Similarity: 80.97% Conservative: 51  
Best local Similarity: 72.74% Mismatches: 98  
Query Match: 51.59% Indels: 42  
DB: 14 Gaps: 15  
US-09-932-678-2 (1-651) x YI6973 (1-1800)  
QY 24 LeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhePheAsn 43  
Db 26 CTGAGTGCCTCGCGACTGGCTTCAAACATGCTTGCAATTAGAGAGTGATTC----- 79  
QY 44 SerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyr 63  
Db 80 TCAGTCTCCCAAGAAACTGTCGTTGG--TGGACTGTGACAGACGCTTTACTGAGT--- 133  
QY 64 LysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIle 83  
Db 134 ACAAAAGGGTGAACAAATGACTTAGAGTTGTTGAAGAACAG-CTGCTGATCCTGATATA 192  
QY 84 LysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThr 103  
Db 193 GAGGATGATCAGATCATTAATACTGGCTGCTGGAATTCGTTCTCTGTCATGTAATTGACA 252  
QY 104 LysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGln 123  
Db 253 AAAGACTTTGAGCAACTTATAAACATCATATGAGATTGCCCTGGTTGAATAGAGTCAG 312  
QY 124 ThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPhe 143  
Db 313 ACGGTGGTGGAGGAGTATTTGGCTTTTCTTGGTAATCTTGTGTGACGACAGACTGCTTTC 372  
QY 144 LeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleLys 163  
Db 373 CTTAGACCATGCTCTCAGCATGATTGCTTCTCATTTTGTACTTCCCCGAGTATTGTCAAG 432  
QY 164 GluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPhe 183  
Db 433 GAAGGTGGCATTGATGT-TCAGAGTCTGAGGATGAAGATGACAATCTTCTCGCAAATTTT 491  
QY 184 AspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPhe 203  
Db 492 GACACATGTCACAAAGCCTTGCAATAATAACAAGATATGTCCTTCGACACCATGGTTT 551  
QY 204 LeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGlu 223  
Db 552 CTAATGCCAATACTGGTAGAAAAAATTTCCATTTGTGAGGAAATCAGAGAGAACACTGGAA 611  
QY 224 CysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIle 243  
Db 612 TGCTATGTTCATAACTATTAAAGGATAAGTTTATATTTCCCAACGTTGAGCGGTGAAATT 671  
QY 244 LeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIle 263  
Db 672 CTCGAGCTTATTATTGAAGTCA-CTCAAGTTAGATGTGAGTGTGTCCGCGCAGGATATT 730  
QY 264 GluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeu 283  
:::|||||

/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="IRL573"  
/tissue\_type="liver"  
/dev\_stage="adult"  
/clone\_lib="Rat liver ESTs (E.Olivier)"  
/note="Organ: liver; Vector: lambda ZAPII; Library  
construction: Ripperger,J.A. et al. J. Biol. Chem. (1995)  
270(50):29998-30006"



Db 731 GACGATGCTGAAGACAGCAGCTCAGACTTATGGTGGACAGACGCCGAGAGGACTG 790

Qy 284 PheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArgLeu 303

Db 791 TT-AATATGGATGA-GATGAGGA-ACTGACCTTGAGAAAGGCTGAGCAGGCGCCTG 847

Qy 304 AspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSer 323

Db 848 GC-CGGATGGCTCATCCACTGCAGAGCGCTGGACATCTGCTGTCTGCTCTCC 906

Qy 324 TyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeu 343

Db 907 TACATTGACGATGTCTGCCATGTGC--AGTAAATTTGATAACAATAAAACAAAGA--TTA 962

Qy 344 TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHis 363

Db 963 TATCGTATCTGATATCTATCTTG--ACAACGTGTGTGCCACACATGCCTC-TGCCAT 1018

Qy 364 ValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeu 383

Db 1019 GTACAGTTCTTTCATGTTTTCCTTGACGT---CAAATGGGGTTTGCAGAGCATTTTGT 1075

Qy 384 GluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAla 403

Db 1076 GAACATCTC---TGAAAAGTGCAGGATCA-AATAACCCCTGTCTATCATCAGCAGCTGCTGC 1131

Qy 404 GlyAsnTrpIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLys 423

Db 1132 ---AAATACATG---AGCTTTTGGC-AGAACTAAATTAAT-CCTCTTATCACTGTGAAG 1183

Qy 424 SerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGly 443

Db 1184 TCATGCCTGGATCTTTGGGT-AACTGGCTGCACTG-TACCTTAATAGCCAG-----GTT 1235

Qy 444 ThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaVal 463

Db 1236 CGGACAAAGGCTTTGTATGTGC--ACTCATGGACCATTATCAGC---TGCCAAAGCTGTG 1290

Qy 464 PheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeu 483

Db 1291 TC-TACACT---GTGTTTAGACACAGCACTTTGAGTGCACCT-----GAGGAAGT-CTA 1339

Qy 484 GlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLys 503

Db 1340 CGGTATCTCCAGAGTCTAAATTTTGAACGCATGTGTGAGCCAGCTGAACCCACTGAAG 1399

Qy 504 IleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuVal 523

Db 1400 ATCTGCCTGCCACGTGTGTATCTCTTGTCTGCATACAATA-----ATACAGCTGTGT 1450

Qy 524 PheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThr 543

Db 1451 CTGCTACAC---ATCATGGAGAGGAC-AATCGCCAGATGTCTCCCTGTTATCG--AGTACA 1504

Qy 544 AlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePropheAsp 563

Db 1505 GCTGGTGGAGACTCTGTGCAGACCTGCACCAACCCATTGGACACTTTCTTCCCTTTGAC 1564

Qy 564 ProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAsp 583

Db 1565 CCTTGATGTCTAAAGAGGTCAAAGAAGTTTCATCGATCTCTGCTATCAGGCTTGGGAAGAC 1624

Qy 584 MetSerAlaGluGluLeuGlnPheLysLysProMetLysLysAspIleValGluAsp 603

Db 1625 ATGAGTGATGAAGAGCTGCAGGAGCTTAGGAACCCACT-AGAAAGGAGCTAGTGGAGGAT 1683

Qy 604 GluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThr 623

Db 1684 GAAAGACGATGACTTTTGAAGCGGAGTGC---CAGAGTGACACAGTGGCTTGCCTTACC 1740

Qy 624 ProSerSerPheAspThrHisPheArgSerProSerSerValGlySerProProVal 643

Db 1741 CCGAACTCCTTTTGACACCCACTTTCGAGT-CCTTCCAGTAGTGTGGGCTCCCTCCTGTG 1799

RESULT 4

BQ214727

LOCUS BQ214727

DEFINITION AGENCOURT\_7595062 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6062849 5', mRNA sequence.

ACCESSION BQ214727

VERSION BQ214727.1 GI:20396127

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 896)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLAM1334 row: p column: 18  
High quality sequence stop: 595.  
Location/Qualifiers

FEATURES

source

1..896

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6062849"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_72"

/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."

ORIGIN

Alignment Scores:

Pred. No.: 3.41e-144 Length: 896

Score: 1290.00 Matches: 265

Percent Similarity: 91.22% Conservative: 5

Best Local Similarity: 89.53% Mismatches: 13

Query Match: 37.91% Indels: 13

DB: 13 Gaps: 4

US-09-932-678-2 (1-651) x BQ214727 (1-896)

Qy 314 LeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGly 333

Db 23 GTGGACATCTGATGTCCTTGGTTTGTCTCTACATGAAGGATGCTGCTATGTAGATGGT 82

Qy 334 LysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLys 353

Db 83 AAGTTTGATAACGGCAAAACAAAGGATCTATATCGGACCTGATAAACATCTTGACAAA 142

Qy 354 LeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSer 373

Db 143 CTCCTGTTGCCACCCATGCTCTGCCATGTACAGTTTTCATGTTTACCTCTGTAGT 202

Qy 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393

Db 203 TTCAAATTGGGATTCGAGAGGCATTTTTCGAAACATCTCTGAAAAAATTCGAGGACCCA 262

Qy 394 SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413

Db 263 AGTAATCCTGCCATCATCAGCAGGCTGCTGGAAATATATTGGAAAGCTTTTGGCAAGA 322





Db 721 GTGGGAAGAAATGAAGAAATGATGAACCTTTCCTGAAAGGGGAAAGTGCCCCCAAGAAAGG 780

QY 617 -----AspThrValIleGlyLeThrProSerSerPheAspThr--HisPhe---A 632

Db 781 AAACCCGTTGATTGTGAATTCACACCCAAACCTCCCTTTTGTGACACGGCCATTTCCCA 840

QY 632 rgSerProSerSerSer-----ValGlySerProProValLeuTyrMetGlnProSerP 650

Db 841 GAAAGCCCTTCTTCTAAATAAGTGGTGGGGCCCTCCC-----CCACACC 885

QY 650 roLeu 651

Db 886 CCGTG 890

RESULT 6

BU510623

LOCUS

DEFINITION BU510623 889 bp mRNA linear EST 12-SEP-2002

AGENCOURT 10120114 NIH\_MGC 134 Mus musculus cDNA clone

IMAGE:6505326 5', mRNA sequence.

ACCESSION BU510623

VERSION BU510623.1 GI:22816856

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 889)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

JOURNAL NIH-MGC <http://mgc.nci.nih.gov/>

COMMENT National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. David Rowe

cdNA Library Preparation: Invitrogen Corp

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

plate: LLAM14066 row: a column: 07

High quality sequence start: 24

High quality sequence stop: 646.

FEATURES

source

1..889

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:6505326"

/tissue\_type="undifferentiated limb"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 134"

/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 4.12e-136 Length: 889

Score: 1223.00 Matches: 244

Percent Similarity: 88.93% Conservative: 13

Best Local Similarity: 84.43% Mismatches: 28

Query Match: 35.94% Indels: 4

DB: 13 Gaps: 1

US-09-932-678-2 (1-651) x BU510623 (1-889)

QY 4 ProLeuLeuHisThrArgLeuProGlyAspAlaAlaAsrSerSerAlaValLysLys 23

Db 15 CCCGCTGCTCCACTTCTTCGGGCGATGTGACGGCAGCGCCTCTGCCACG-----68

QY 24 LeuGlyAlaSerArgThrGlyLeSerAsnMetArgAlaLeuGluAsnAspPhePheAsn 43

Db 69 CTGAGTCCGTCCGGGACTGGGCTTTCGGATATGCTCGGTTAGAGAGCGATTTCCTCAAT 128

QY 44 SerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyr 63

Db 129 TCTCTCCGAAAAAACTGTTCCGTTTGGCGGAAGTGTGACAGAAGTCTTACTGAATAC 188

QY 64 LysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIle 83

Db 189 AAAAAAGGTGAACAAATGACTTAGAGTTGTTGAAGAACCCAGCTGTCTGTATCCTGATATA 248

QY 84 LysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThr 103

Db 249 AAGGATGACCAGATCACTTAAGTGGCTACTAGAAATTCGTTCTCTCATGTACTTGACA 308

QY 104 LysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGln 123

Db 309 AAAGACTTTGAGCAACTTATAAACATCATATTGAGATTGCCCTGGTTGAATAGAGTCAG 368

QY 124 ThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPhe 143

Db 369 AGAGTGGTGGAGGAGTATTTGGCTTTCTTGGTAATCTTGTGTCTGCACAGACTGTCTTC 428

QY 144 LeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleLys 163

Db 429 CTTAGACCATGTCTCAGCATGATTGCGTCTCATTTTGTACCTCCCCGAGTAATTGTCAAG 488

QY 164 GluGlyAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPhe 183

Db 489 GAAGGTGGCATAGATGTTTCAGATTCTGATGACGAAGATGACAAATCTCTCGCAATTTT 548

QY 184 AspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPhe 203

Db 549 GACACATGTCACAGAGCCTTGCAATAATAACAAGATATGTCCTCATCGACCATGGTTT 608

QY 204 LeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGlu 223

Db 609 CTAATGCCAATATCTGTTGGAGAAATTCCTGTTTGTGAGGAAGTCCGAGAGAACATTGGAA 668

QY 224 CysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIle 243

Db 669 TGTWATGTTTCATAACTATTAAAGGATAAGTTTATATTTCCTCCCACTTTGAGCGTGAAT 728

QY 244 LeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIle 263

Db 729 CTGGAGCTTGNTATTGAAAAGCTACTCAAGTTAGATGTAGTGTATCCGGCAGGATATT 788

QY 264 GluAspAlaGlu-GluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLe 283

Db 789 GAAGATGCTGAAAGAGAAAGCAGCTCAAACTTGTGTGGGACAGATACCCCGGAAAGACT 848

QY 283 u-PheAsnMetAspGluAspGlu 290

Db 849 GGTTTATTATGGATGAAGATGAA 871

RESULT 7

BU610641

LOCUS

DEFINITION BU610641 828 bp mRNA linear EST 20-FEB-2003

UI-M-FC0-caq-n-01-0-UI.r1 NIH\_BMAP\_FC0 Mus musculus cDNA clone

UI-M-FC0-caq-n-01-0-UI 5', mRNA sequence.

ACCESSION BU610641

VERSION BU610641.1 GI:23276856

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 828)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. James Lin, Univeristy of Iowa



contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

ORIGIN

Alignment Scores:  
Pred. No.: 2.54e-134 Length: 788  
Score: 1207.50 Matches: 245  
Percent Similarity: 94.34% Conservative: 5  
Best Local Similarity: 92.45% Mismatches: 10  
Query Match: 35.48% Indels: 6  
DB: 12 Gaps: 1

US-09-932-678-2 (1-651) x BG502871 (1-788)

Qy 302 ArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuVal 321  
Db 2 CGGCTGACCATGGTGATCTCTGTAGCCGAGCGCCTGGACATCCTGATGCTTTGGTT 61  
Qy 322 LeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLys 341  
Db 62 TTGTCTTACATGAAGGATGCTGCTATGTAGATGGTAAGGTGATAACGGCAAAACAAG 121  
Qy 342 AspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSer 361  
Db 122 GATCTATATCGGACCTGATAACATCTTTGACAACTCTCTGTTGCCACCCATGCCTCC 181  
Qy 362 CysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAla 381  
Db 182 TGCCATGTACAGTTTTTCATGTTTTTACCTCTGTAGTTTCAAAATGGGATTCGACAGGCA 241  
Qy 382 PheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGln 401  
Db 242 TTTTGGAAACATCTCTGGAAAAAATTGCAGGACCCCAAGTAATCCTGCCATCATCAGGCAG 301  
Qy 402 AlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThr 421  
Db 302 GCTGCTGGAAATTATATTGGAAAGCTTTNTGGCAAGAGCTAAATTTATTCTCTTATTACT 361  
Qy 422 ValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAsp 441  
Db 362 GTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAGGAT 421  
Qy 442 SerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGln 461  
Db 422 TCGGGAACAAGGCATTTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAA 481  
Qy 462 AlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGlu 481  
Db 482 GCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAA 541  
Qy 482 GlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501  
Db 542 GGTNGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTATGAGCCAGCTAAATCCC 601  
Qy 502 LeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsn-LysTyrG1 521  
Db 602 CTGAAGATTTGCTG-CCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATTAAGTACCA 660  
Qy 521 nLeuValPheCysTyrThrIleIle-GluArgAsnAsnArgGln-MetLeuProVal-I1 540  
Db 661 GCTCGTCTTCTGTTACACCATCATTTGGAGAGGAACAATCGGCAGAATGGTTGCAGTCCAT 720  
Qy 540 eArgSerThrAlaGlyGlyAsp---SerValGlnIleCysThrAsnProLeuAspThrPh 559  
Db 721 TGGGAGTACCGCTGGAGGAGAACTCAGTGCCGAATCTGGGCAAAACCCGGTGGAAACCTT 780  
Qy 559 ePhePro 561  
Db 781 TTTTCCC 787

RESULT 9

BF203684 996 bp mRNA linear EST 06-NOV-2000  
LOCUS 601866331F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4099398 5',  
DEFINITION mRNA sequence.  
ACCESSION BF203684  
VERSION BF203684.1 GI:11097270  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 996)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCW967 row: n column: 07  
High quality sequence stop: 684.  
Location/Qualifiers  
1. .996  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4099398"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_17"  
/note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;  
Site 2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.88e-124 Length: 996  
Score: 1125.50 Matches: 238  
Percent Similarity: 88.24% Conservative: 2  
Best Local Similarity: 87.50% Mismatches: 26  
Query Match: 33.07% Indels: 9  
DB: 10 Gaps: 2  
US-09-932-678-2 (1-651) x BF203684 (1-996)  
Qy 379 AlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIle 398  
Db 3 GCAGAGGCATTTTGGACATCTCTGGAATAATG-CAGGACCCCAAGTAATCCTGCCATC 61  
Qy 399 IleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIlePro 418  
Db 62 ATCAGGCAGGCTGCTGGAAATATATATTGGAAGCTTTTCGGCAAGAGCTAAATTTATTCCT 121  
Qy 419 LeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsn 438  
Db 122 CTTATTACTGTAAATCATCGCCTAGATCTTGTGGTTAACTGGCTGCACATATACCTTAAT 181  
Qy 439 AsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSer 458  
Db 182 AACCAGGATTCGGGAACAAAGGCATTCGCGATGTTGCTCTCCATGGACCATTTTACTCA 241  
Qy 459 AlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsn 478  
Db 242 GCCTGCCAAGCTGTGTTCTACACCTTTGTGTTTAGACACAAGCAGCTTTTGAGCGGAAAC 301

```
QY      479  LeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGln 498
      |||
Db      302  CTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTATGAGCCAG 361
      |||
QY      499  LeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsn 518
      |||
Db      362  CTAATCCCTGAAGATTTGCTGCCCTCAGTGGTTAACTTTTGTGCTGCAATCACAAAT 421
      |||
QY      519  LysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuPro 538
      |||
Db      422  AAGTACCAGCTCGTCTTCTGTACACCATCATGTGAGAGGACAATCGCCAGATGCTGCCA 481
      |||
QY      539  ValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAsp-Th 558
      |||
Db      482  GTCATTAGGAGTACCGCTGGAGGAGACTCAGTCAGATCTG-ACAAACCCGCTGGACAAC 540
      |||
QY      558  rPhePheProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTy 578
      |||
Db      541  CTTCTTCCCTTTGATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTA 600
      |||
QY      578  rGlnValTrpGluAspMetSerAlaGluGlnLeuGlnPheLysLysProMetLysLy 598
      |||
Db      601  TCAGGTGTGGGAAGACATGATGCTGAAGAGCTACAGGAGTCAAGAAACCATGAAAAA 660
      |||
QY      598  sAspIleValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspTh 618
      |||
Db      661  GGACATAGTGAAGATGAAGATGATGC-ATTCTGAAAGGAGATGGCCAGATGGA----- 714
      |||
QY      618  rValIleGlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerVa 638
      |||
Db      715  ----TACGGGATGGGATTAGACAGCTCTTGTGAACAGATTTTCAAGTCTACAAAAG----- 765
      |||
QY      638  lGlySerProProValLeuTyrMetGlnProSer 649
      |||
Db      766  -GGAGAGACCCACCAGGATGAATGACACCAAGT 798
      |||
RESULT 10
CA511975
LOCUS      754 bp mRNA linear EST 15-NOV-2002
DEFINITION UI-R-FJ0-cpx-i-24-0-UI.r1 UI-R-FJ0 Rattus norvegicus cDNA clone
            UI-R-FJ0-cpx-i-24-0-UI 5', mRNA sequence.
ACCESSION  CA511975
VERSION     CA511975.1 GI:25002929
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 754)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. James Lin, Universtiy of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE.
FEATURES   Location/Qualifiers
            source          1..754
```

```
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-FJ0-cpx-i-24-0-UI"
/tissue_type="embryo"
/dev_stage="embryo"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FJ0"
/note="Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I;
UI-R-FJ0 is a cDNA library containing the following
tissue(s): rat embryo. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pTT3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CATCTCTACT. This library
was created for the University of Iowa Program for Rat
Gene Discovery and Mapping (Val Sheffield, Bento Soares
and Tom Casavant)"
```

ORIGIN

```
Alignment Scores:
Pred. No.:      8.66e-124      Length:      754
Score:          1120.00      Matches:      226
Percent Similarity: 51.39%      Conservative: 15
Best Local Similarity: 48.19%      Mismatches: 10
Query Match:      32.91%      Indels:      218
DB:              14          Gaps:          1

US-09-932-678-2 (1-651) x CA511975 (1-754)

QY      139  AlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValPro 158
      |||
Db      1      GCACAGACTGTCTTCCTTAGACCATGTCTCAGCATGATTGCTTCTCATTTTGTACCTCC 60
      |||
QY      159  ArgValIleIleLysGluGlyAspValAspValSerAspSerAspGluAspAspAsn 178
      |||
Db      61  CGAGTTATTGTCAAGGAAGGTGGCATTGATGTTTCAGAGTCTGAGGATGAAGATGACAA 120
      |||
QY      179  LeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValPro 198
      |||
Db      121  CTTCTCTGCAAAATTTTGACACATGTCAAAAGCCTTGCAATAATAACAAGATATGTCCT 180
      |||
QY      199  SerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSer 218
      |||
Db      181  TCGACACCATGGTTTCTAATGCCAATA----- 207
      |||
QY      219  GluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThr 238
      |||
Db      207  ----- 207
      |||
QY      239  LeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAla 258
      |||
Db      207  ----- 207
      |||
QY      259  SerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAsp 278
      |||
Db      207  ----- 207
      |||
QY      279  SerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAla 298
      |||
Db      207  ----- 207
      |||
QY      299  GlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMet 318
      |||
Db      207  ----- 207
      |||
QY      319  SerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGly 338
      |||
```



Db 207 ----- 207

Qy 339 LysThrIysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThr 358

Db 207 ----- 207

Qy 359 HisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPhe 378

Db 207 ----- 207

Qy 379 AlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIle 398

Db 207 ----- 207

Qy 399 IleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIlePro 418

Db 207 ----- 207

Qy 419 LeuIleThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsn 438

Db 208 -----CTGGATCTTTTGGTTAACTGGCTGCACCTGTACCTTAAT 246

Qy 439 AsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPropheTyrSer 458

Db 247 AGCCAGGATTCGGGAACAAAGGCTTTTGTGATGTTGCACTCCATGGACCATTTTATTCA 306

Qy 459 AlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsn 478

Db 307 GCCTGCCAAGTGTTCTACACTGTTGTTTGTAGACACAAAGCAACTTTTGAGTGGCAAC 366

Qy 479 LeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGln 498

Db 367 CTGAAGGAAGGTCTACGGTATCTTCAGAGTCTAAATTTTGAACGCATTGCTGAGCCAG 426

Qy 499 LeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsn 518

Db 427 CTGAACCCACTGAAGATCTGCTGCCACGCTGGTTAAITTTCTTGCTGCCATAACAAAT 486

Qy 519 LysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuPro 538

Db 487 AAATACCAGCTGTGTTCTGTACACCATCATGGAGAGGAACAATCGCCAGATGCTCCCT 546

Qy 539 ValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThr 558

Db 547 GTTATTCCGAAGTACAGCTGGTGGAGACTCTGTGCAGACCTGCACCAACCCATTGGACACT 606

Qy 559 PhePheProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyr 578

Db 607 TTCTTTCCCTTTGACCCCTTGATGCTTAAGAGGTCAAAGAAGTTCATCGATCCTGTCTAT 666

Qy 579 GlnValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLys 598

Db 667 CAGGCTTGGGAAGACATGATGATGAAGAGCTGCAGGAGCTTANGAAACCCACTAGAAAG 726

Qy 599 AspIleValGluAspGluAspAsp 607

Db 727 GAGCTAGTCGAGGATGAAGACGATGAC 753

RESULT 11

CF738084

LOCUS

DEFINITION UI-M-HD0-ckt-1-24-0-UI.r1 NIH\_BMAP\_HD0 Mus musculus cDNA clone

IMAGE:30611903 5', mRNA sequence.

CF738084

CF738084.1 GI:37634421

EST.

Mus musculus (house mouse)

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 769)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE  
JOURNAL  
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.

FEATURES  
source

Location/Qualifiers  
1..769  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30611903"  
/tissue\_type="whole eye"  
/dev\_stage="embryo 12.5,13.5,14.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP HD0"  
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is TTATTGAAGT. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:  
Pred. No.: 2.36e-123 Length: 769  
Score: 1116.50 Matches: 212  
Percent Similarity: 90.62% Conservative: 20  
Best Local Similarity: 82.81% Mismatches: 23  
Query Match: 32.81% Indels: 1  
DB: 14 Gaps: 1

US-09-932-678-2 (1-651) x CF738084 (1-769)

Qy 148 LeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVal 167  
Db 3 CTCAGCATGATTGCGTCTCATTTTGTACTCTCCCGAGTAATTGTCAAGGAAGGTGGCATA 62  
Qy 168 AspValSerAspSerAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHis 187  
Db 63 GATGTTTCAGATTCTGATGACGAAGATGACAATCTTCTGCAATTTTTCACACATGTTCAC 122  
Qy 188 ArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIle 207  
Db 123 AGAGCCTTGCAATAATAACAAGATATGTCCCATCGACACCATGGTTTCTAATGCCAATA 182  
Qy 208 LeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHis 227  
Db 183 CTGGTGGAGAAATTCCCGTTTGTGAGGAAGTCCGAGAGAACATTTGGAATGTTATGTTTCAT 242  
Qy 228 AsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIle 247  
Db 243 AACTTATTAAGGATAAGTTATATTATTTCCCACTTTTGAGCGTGAAATCTGGAGCTTGT 302  
Qy 248 IleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGlu 267  
Db 303 ATTGAAAAGCTACTCAAGTTAGATGTGAGTGTATCGCGCAGGATATTGAAGATGCTGAA 362





VERSION BI561523.1 GI:15448837  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 733)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM11755 row: 1 column: 06  
High quality sequence stop: 717.  
FEATURES  
source  
1. .733  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5298581"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_97"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
(gtcgag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTNN-3', size-selected for average  
insert size 2.2 kb and normalized to ROT 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHGRI, National Institutes of Health). Note: this is  
a NIH\_MGC Library."  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.36e-122 Length: 733  
Score: 1105.00 Matches: 232  
Percent Similarity: 97.48% Conservative: 0  
Best Local Similarity: 97.48% Mismatches: 4  
Query Match: 32.47% Indels: 5  
DB: 12 Gaps: 0  
US-09-932-678-2 (1-651) x BI561523 (1-733)  
QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerAla 20  
Db ATGGCGGCACCGCTGCTTCACACGCGTTTCGCCGGAGATGCGGCCGCTTCCTCTGCA 82  
QY 21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
Db GTTAAGAAGCTGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGTAGAGAAATGAC 142  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
Db TTTTTCAAATCTCCCCCAAGAAACAACTGTCGGTTTGGTGAACTGTGACAGAACTCTTG 202  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
Db CTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT 262  
QY 81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
Db CCAGACATAAAGGATGACCAGATCATCAACTGGTTGCTAGAAATTCGGTTCCTATCATG 322  
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn 120

Db 323 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTGGTTGAAT 382  
QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
Db AGAAGTCAAAACAGTAGTGAAGAGTATTGGCTTTTCTTGTAATCTTGTATCAGCACAG 442  
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
Db ACTGTTTCTCTCAGACCGTGTCTCAGCATGATTGCTTCCATTTTGTGCTTCCCGAGTG 502  
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180  
Db ATCATTAGGAAGCGATGTAGATGTTTCAGATTCTGTGATGAAGATGATAATCTTCCT 562  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
Db GCAAA-TTTGACACATGTCACAGAGCCCTTGCAATAATAAGCAAGATATGTACCATCGACA 621  
QY 201 ProTrp-PheLeuMetProIleLeu-ValGluLysPheProPheValArgLysSerGluA 220  
Db CCGTGGGTTTCTCATGCCAATACTNGGTGGAAAAAATTTCCATTTGTTCTG-AAATCAGAGA 680  
QY 220 rgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe 236  
Db GAACACTGGAATGTTACGTTCTATAACTTACTAAGGATTAG-GTATATTTC 729  
RESULT 14  
BG428305  
LOCUS BG428305 1037 bp mRNA linear EST 14-MAR-2001  
DEFINITION 602498983F1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4612605 5',  
mRNA sequence.  
ACCESSION BG428305  
VERSION BG428305.1 GI:13334811  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1037)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM1361 row: e column: 22  
High quality sequence stop: 637.  
FEATURES  
source  
1. .1037  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4612605"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_75"  
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site 1:  
SfiI (ggcgccctcgcc); Site 2: SfiI (ggccattatggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.65  
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH\_MGC Library."  
ORIGIN

Alignment Scores: 3.19e-121 Length: 1037  
Pred. No.: 1100.50 Matches: 252  
Score: 89.00% Conservative: 15  
Percent Similarity: 84.00% Mismatches: 19  
Best Local Similarity: 32.34% Indels: 15  
Query Match: 12 Gaps: 3  
DB:

US-09-932-678-2 (1-651) x BG428305 (1-1037)

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerSerAla 20  
Db 22 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGAGATCGCGCGCTTCGTCTCGCA 81

QY 21 VallLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
Db 82 GTTAAGAAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGAAATGAC 141

QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
Db 142 TTTTTCAAATTCCTCCCCAAGAAAACTGTTCCGTTTGGTGGAAGTGTGACAGAAGTCTTG 201

QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
Db 202 CTGAAGTACAAAAGGGTGAACAAATGACTTTTCAAGTTGTTGAAGAACCCAGCTGTTAGAT 261

QY 81 ProAspIleLysAspAspGlnIlelleAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
Db 262 CCAGACATAAAGGATGACCAGATCATCAACTGGGTGCTAGAAATCCGTTCTTCTATCATG 321

QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIlelleLeuArgLeuProTrpLeuAsn 120  
Db 322 TACTTGACAAAAGACTTTTGAGCAACTTATCAGTATATATTAAGATTGCCTTGGTTGAAT 381

QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
Db 382 AGAAGTCAAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAG 441

QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArg-Va 160  
Db 442 ACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTCCCGGAAGT 501

QY 160 lIlelleLysGluGlyAspValAspValSerAspSerAspGluAspAspAsnLeuPr 180  
Db 502 GATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCC 561

QY 180 oAlaAsnPheAspThrCysHisArgAlaLeuGlnIlelleAlaArgTyrValProSerTh 200  
Db 562 TGCAAA-TTTGACACATGTACAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGAC 620

QY 200 rPro-TrpPheLeuMetProIleLeu-Val-GluLysPheProPhe-ValArgLysSer- 218  
Db 621 ACCGTTGGTTTCTCATGCCAATACTGGGTGGCAACAATTCCTCATATGTTGCAAAAACTC 680

QY 219 --GluArgThrLeuGlu-CysTyr-ValHis-AsnLeuLeu-ArgIleSerValTyrPhe 236  
Db 681 CAGAGAGAACAACACTGGAACCTGTCCCGTTCAATGAACCTTACTTCAGGATCACGGGATATGTC 740

QY 237 ProThrLeu-ArgHis-GluIleLeuGluLeuIlelleGluLysLeuLeuLysLeuAspV 256  
Db 741 CCAAGCTTGGAGGCATGCAAAATTCGGAGCTTATATTGAAAACCTACTCAAGGTGGACTG 800

QY 256 alAsnAlaSerArgGlnGlyIle--GluAspAla--GluGluThrAlaThrGlnThrC 274  
Db 801 TGAATGGCTCCCGGGCGGGGTATTGGAAGACTGCTGAGGAACGACGGAAGTACAAAACCTG 860

QY 274 ysGlyGlyThrAspSerThrGluGlyLeuPheAsnMet 286  
Db 861 GTGGGGGAACAATAATCCCGGCAAGGATTTGTTCACTTGTG 898

RESULT 15  
BE734354

LOCUS BE734354 757 bp mRNA linear EST 15-SEP-2000  
DEFINITION 601565596F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3840623 5',  
mRNA sequence.  
ACCESSION BE734354  
VERSION BE734354.1 GI:10148346  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 757)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM530 row: o column: 24  
High quality sequence stop: 693.  
Location/Qualifiers  
1. 757  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3840623"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC\_21"  
/note="Organ: placenta; Vector: pOTF7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:  
Pred. No.: 1.2e-120 Length: 757  
Score: 1094.00 Matches: 225  
Percent Similarity: 93.63% Conservative: 10  
Best Local Similarity: 89.64% Mismatches: 12  
Query Match: 32.15% Indels: 6  
DB: 10 Gaps: 0

US-09-932-678-2 (1-651) x BE734354 (1-757)

QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIlelleAlaArgTyrValProSerThr 200  
Db 3 GCAAATTTTGACACATGTACAGAGCCTTGCAACTAATAGCAAGATATGTACCATCGACA 62

QY 201 ProTrp-PheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluAr 220  
Db 63 CCGTGTATTTCTCATGCCAAATCTGGTGGAAAAATTTCCATTGTTTCGAAAAATCAGAGAG 122

QY 220 gThrLeuGlu-CysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuA 240  
Db 123 AACACTGGACATGTTACGTTTCATAACTTACTTAAGGATTAGTGATATATTTTCCAACCTTGA 182

QY 240 rgHisGluIleLeuGluLeuIlelleGluLysLeuLeuLysLeuAspValAsnAlaSerA 260  
Db 183 GGCATGAAATTTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGGATGTGAATGCATCCC 242

QY 260 rgGlnGly-IleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSer 279  
Db 243 GGCAGGGCTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCC 302

QY 280 ThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLys-AlaGl 299  
Db |||||  
303 ACGGAAGGATTGTTTAAATATGGATGAAGATGAAGAAACTGAACATGAACAAACGGCTGG 362  
QY 299 YProGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSe 319  
Db |||||  
363 TCCTGAACGGCTCGACCCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGTC 422  
QY 319 rLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLy 339  
Db |||||  
423 TTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTGGATAACGGCAA 482  
QY 339 sThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHi 359  
Db |||||  
483 AACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTAGCCCCAACCA 542  
QY 359 sAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAl 379  
Db |||||  
543 TGCCCTCCTGCCATGTACAGTATTACATGTATTACCTCTGTAGTTTCAAATCGGGATTGCG 602  
QY 379 aGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIl 399  
Db |||||  
603 AGAGGCATATTAGGAACATCTCTGGAAAAAATTCAGGACCCCAAGTAATCCTGCCATCAT 662  
QY 399 eArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLe 419  
Db |||||  
663 CAGGCAGGCTGCTGGAAATTATATGGGAAGCTT-ATGGGAAGAGCCCTAATTAATTCCTCT 721  
QY 419 uIleThrValLysSerCysLeuAsp 427  
Db |||||  
722 ATTACTGGTCCA-TCATGCCCTAGAA 745

Search completed: July 18, 2004, 12:07:14  
Job time : 4426 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 01:15:01 ; Search time 8111 Seconds  
(without alignments)  
11050.844 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagaggctgtggctggaag.....tgacatttgggattcccat 2068

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2068	100.0	2068	9	AF227156	AF227156 Homo sapi
2	2020	97.7	3756	9	HSA272050	AJ272050 Homo sapi
3	2018.4	97.6	2040	6	AX036047	AX036047 Sequence
4	2008	97.1	3709	9	BC036182	BC036182 Homo sapi
5	1211.6	58.6	135873	2	AC012547	AC012547 Homo sapi
6	1211.6	58.6	192815	9	AC013444	AC013444 Homo sapi
7	1179.8	57.1	3166	10	BC034110	BC034110 Mus muscu
8	894	43.2	1423	6	AX427479	AX427479 Sequence
9	567	27.4	2493	6	AX713860	AX713860 Sequence
10	567	27.4	2493	9	AK055742	AK055742 Homo sapi
11	506.4	24.5	1168	9	BC009198	BC009198 Homo sapi
12	478	23.1	2243	10	BC055781	BC055781 Mus muscu
13	420.8	20.3	1461	6	AX274907	AX274907 Sequence
14	420.8	20.3	2627	9	BC006441	BC006441 Homo sapi
15	420.8	20.3	3645	9	AK126166	AK126166 Homo sapi
16	420	20.3	437	6	BD073832	BD073832 5'EST of
17	292	14.1	341	6	BD072397	BD072397 Secreted
18	208.2	10.1	513	6	AR415449	AR415449 Sequence
19	208.2	10.1	513	6	BD111002	BD111002 EST and e
20	208	10.1	129473	2	AC040158	AC040158 Homo sapi
21	208	10.1	143783	9	AC139256	AC139256 Homo sapi
22	208	10.1	184471	9	AC136026	AC136026 Homo sapi
23	208	10.1	184891	9	AC092137	AC092137 Homo sapi
24	208	10.1	189770	2	AC139281	AC139281 Homo sapi
25	208	10.1	194670	2	AC141613	AC141613 Homo sapi
26	208	10.1	196687	2	AC136442	AC136442 Homo sapi
27	208	10.1	215222	2	AC141467	AC141467 Homo sapi
28	208	10.1	215974	2	AC141607	AC141607 Homo sapi
29	208	10.1	218593	2	AC141303	AC141303 Homo sapi
30	208	10.1	241278	9	AC097268	AC097268 Pan trogl
31	203.6	9.8	213	6	AX907303	AX907303 Sequence
32	203.6	9.8	213	6	BD042836	BD042836 Sequence
33	192	9.3	131839	9	AC112166	AC112166 Homo sapi
34	192	9.3	175691	2	AC007615	AC007615 Homo sapi
35	192	9.3	191496	2	AC009130	AC009130 Homo sapi
36	192	9.3	207997	9	AC009093	AC009093 Homo sapi
37	190.8	9.2	183586	9	AC092562	AC092562 Papio ham
38	189.8	9.2	218593	2	AC141303	AC141303 Homo sapi
39	188.8	9.1	173275	9	AC092375	AC092375 Homo sapi
40	188.8	9.1	179150	2	AC017077	AC017077 Homo sapi
41	188.8	9.1	191598	2	AC144362	AC144362 Gorilla g
42	188.8	9.1	202004	9	HUAF001549	AF001549 Human Chr
43	188.8	9.1	221030	2	AC145403	AC145403 Gorilla g
44	187.2	9.1	157631	2	AC142079	AC142079 Homo sapi
45	187.2	9.1	170613	9	AC106788	AC106788 Homo sapi

ALIGNMENTS

RESULT 1  
AF227156  
LOCUS AF227156 2068 bp mRNA linear PRI 29-APR-2000  
DEFINITION Homo sapiens RRN3 mRNA, complete cds.  
ACCESSION AF227156  
VERSION AF227156.1 GI:7670099  
KEYWORDS .  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2068)  
AUTHORS Moorefield,B., Greene,E.A. and Reeder,R.H.  
TITLE RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)  
MEDLINE 20243763  
PUBMED 10758157  
REFERENCE 2 (bases 1 to 2068)  
AUTHORS Moorefield,B., Greene,E.A. and Reeder,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer  
Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA  
FEATURES Location/Qualifiers  
source 1..2068  
    /organism="Homo sapiens"  
    /mol\_type="mRNA"  
    /db\_xref="taxon:9606"  
    /chromosome="16"  
    /map="16p12"  
    69..2024  
    /note="similar to yeast RNA polymerase I transcription  
    factor Rn3"  
    /codon\_start=1  
    /product="Rn3"  
    /protein\_id="AAF66160.1"  
    /db\_xref="GI:7670100"  
    /translation="MAAPLLHTRLPGDAAASSAVKKLGAASRTGTGISMRALENDFENS  
    PRTKTVRFSGTVTEVLLKKYKGETNDFELKNQLLDPDKDDQIINWLEFRSSIMYL  
    TKDFEQLISILRLPWLNSQTVVEEYLAFLGNLVSQAQTVFLRPLCLSMIAHFVPPRV  
    IIKEGVDVSDDEDDNLNANFDTCHRALQIIARIYVSTFWFLMPILVEKFPFVRKS  
    ERTLECYVHNLRLISVYFPTLRHEILELIEKLKLLDVNASRQGIEDIAEETATQTCGG  
    TDSTEGLFNWDEDETEHETKAGPERLDQMHPVAERLDILMSLVLSYMKDVCYVDGK  
    VDNGTKLDLYLDINIFDKLLPHTASCHVOFFMFYLCSPKLGFAEAFLEHLWKLLQD  
    PSNPAILRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNLWLIYLNNDQSGTKAFCDV  
    ALHGPFYSACQAVFYTFVRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS  
    VNFFAAITNKYLQVFCYTIIBERNRQMLPVIRSTAGGDSVQICTNPLDTPFPDPCV  
    LKRSKKFIDPIYQVWEDMSAEELQEFKKPKMKDIVEDEDDFLKGEVPPQNDVIGITP  
    SSFDTFRSPSSSVGPPVLYMQPSPL"

ORIGIN

Query Match 100.0%; Score 2068; DB 9; Length 2068;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2068; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTGAGCGCGAGCGGTGCGTTAGT 60  
DB 1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTGAGCGCGAGCGGTGCGTTAGT 60  
QY 61 TCGGCCCCAATGGCGGCACCGCTGCTTCACACGCGTGTTCGCGGAGATGCGCGCGCTTCGT 120  
DB 61 TCGGCCCCAATGGCGGCACCGCTGCTTCACACGCGTGTTCGCGGAGATGCGCGCGCTTCGT 120  
QY 121 CCTCTGCAGTTAAGAAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATGCGTGCATTAG 180  
DB 121 CCTCTGCAGTTAAGAAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATGCGTGCATTAG 180  
QY 181 AGAATGACTTTTCAATTTCCCCCAAGAAAAAAGTGTTCGCTTTGGTGGAACTGTGACAG 240  
DB 181 AGAATGACTTTTCAATTTCCCCCAAGAAAAAAGTGTTCGCTTTGGTGGAACTGTGACAG 240  
QY 241 AAGTCTTGCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCCAGC 300  
DB 241 AAGTCTTGCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCCAGC 300  
QY 301 TGTTAGATCCAGACATAAAGGATGACCAATCAACTGGCTGCTAGCAATCCGTTCTT 360  
DB 301 TGTTAGATCCAGACATAAAGGATGACCAATCAACTGGCTGCTAGCAATCCGTTCTT 360  
QY 361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAGATTGCCTT 420  
DB 361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAGATTGCCTT 420  
QY 421 GGTGAATAGAAAGTCAAAACAGTAGTGAAGATTTGGCTTTCTTGGTAATCTTGTAT 480  
DB 421 GGTGAATAGAAAGTCAAAACAGTAGTGAAGATTTGGCTTTCTTGGTAATCTTGTAT 480  
QY 481 CAGCACAGACTGTTTTCTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTGTGCTC 540

DB 481 CAGCACAGACTGTTTTCTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCTC 540  
QY 541 CCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTTCTGATGATGAAGATGATA 600  
DB 541 CCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTTCTGATGATGAAGATGATA 600  
QY 601 ATCTTCCTGCAAAATTTTGACACATGTACAGAGCGCTTGCAAAATATAGCAAGATATGTAC 660  
DB 601 ATCTTCCTGCAAAATTTTGACACATGTACAGAGCGCTTGCAAAATATAGCAAGATATGTAC 660  
QY 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAAAAATTTCCATTTGTTCCGAAAAAT 720  
DB 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAAAAATTTCCATTTGTTCCGAAAAAT 720  
QY 721 CAGAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGATTAGTGTATATTTTCCAA 780  
DB 721 CAGAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGATTAGTGTATATTTTCCAA 780  
QY 781 CCTTGAGGCATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTGAATG 840  
DB 781 CCTTGAGGCATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTGAATG 840  
QY 841 CATCCCGGAGGTTATTGAAGATGCTGAAAGAAACAGCAAACTCAAACCTTGTGGTGGGACAG 900  
DB 841 CATCCCGGAGGTTATTGAAGATGCTGAAAGAAACAGCAAACTCAAACCTTGTGGTGGGACAG 900  
QY 901 ATTCCACGGAAGGATTGTTTAAATATGATGAAGATGAAGAACTGAACATGAACAAAGG 960  
DB 901 ATTCCACGGAAGGATTGTTTAAATATGATGAAGATGAAGAACTGAACATGAACAAAGG 960  
QY 961 CTGGTCTGAAACGGCTCGACAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCCCTGA 1020  
DB 961 CTGGTCTGAAACGGCTCGACAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCCCTGA 1020  
QY 1021 TGTCTTTTGGTTTTGTCTTACATGAAGGATGTCTGTCTATGTAGATGGTAAAGTTGATAACG 1080  
DB 1021 TGTCTTTTGGTTTTGTCTTACATGAAGGATGTCTGTCTATGTAGATGGTAAAGTTGATAACG 1080  
QY 1081 GCAAAACAAAGGATCTATATCGGACCTGATATAACATCTTTGACAAACTCCTGTTGCCCA 1140  
DB 1081 GCAAAACAAAGGATCTATATCGGACCTGATATAACATCTTTGACAAACTCCTGTTGCCCA 1140  
QY 1141 CCCATGCTCCTGCCATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATTTGGAT 1200  
DB 1141 CCCATGCTCCTGCCATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATTTGGAT 1200  
QY 1201 TCGCAGAGGCATTTTGGAAACATCTCTGAAAAAAATTCAGGACCCCAAGTAATCCTGCCA 1260  
DB 1201 TCGCAGAGGCATTTTGGAAACATCTCTGAAAAAAATTCAGGACCCCAAGTAATCCTGCCA 1260  
QY 1261 TCATCAGGAGGCTGCTGGAATTTATATTGGAAGCTTTTGGCAAGAGCTAAAATTTATTC 1320  
DB 1261 TCATCAGGAGGCTGCTGGAATTTATATTGGAAGCTTTTGGCAAGAGCTAAAATTTATTC 1320  
QY 1321 CTCTTTATTACTGTAAATTCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380  
DB 1321 CTCTTTATTACTGTAAATTCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380  
QY 1381 ATAACCGAGATTTCGGGAACAAAGGCAATTCGCGATGTTGCTCTCCATGGACCATTTTACT 1440  
DB 1381 ATAACCGAGATTTCGGGAACAAAGGCAATTCGCGATGTTGCTCTCCATGGACCATTTTACT 1440  
QY 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGGCGGAA 1500  
DB 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGGCGGAA 1500  
QY 1501 ACCTGAAAGAAGGTTTTCAGATCTTTCAGAGTCTGAATTTTTCAGCGGATAGTAGAGCC 1560  
DB 1501 ACCTGAAAGAAGGTTTTCAGATCTTTCAGAGTCTGAATTTTTCAGCGGATAGTAGAGCC 1560  
QY 1561 AGCTAAATCCCTGAAGATTGCTGCTGCCCTCAGTGGTTAACTTTTGTGCAATCACAA 1620



Db 1561 AGCTAAATCCCCTGAAGATTGGCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCAAA 1620

QY 1621 ATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGC 1680

Db 1621 ATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGC 1680

QY 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAAACCGCTGGACA 1740

Db 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAAACCGCTGGACA 1740

QY 1741 CCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTT 1800

Db 1741 CCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTT 1800

QY 1801 ATCAGGTATGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAAGAAAACCCATGAAA 1860

Db 1801 ATCAGGTATGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAAGAAAACCCATGAAA 1860

QY 1861 AGGACATAGTGAAGATGAAGATGATGACTTTCTGAAAAGCGGAAGTCCCCAGAAATGATA 1920

Db 1861 AGGACATAGTGAAGATGAAGATGATGACTTTCTGAAAAGCGGAAGTCCCCAGAAATGATA 1920

QY 1921 CCGTGATTGGGATCACACCAAGCTCTTTTGACACGCATTTCCGAAGTCTTCAAGTAGTG 1980

Db 1921 CCGTGATTGGGATCACACCAAGCTCTTTTGACACGCATTTCCGAAGTCTTCAAGTAGTG 1980

QY 1981 TGGGCTCCCCACCGTGTGTGATCATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040

Db 1981 TGGGCTCCCCACCGTGTGTGATCATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040

QY 2041 CTGAGATGTGACATTTGGGATTCGCCCAT 2068

Db 2041 CTGAGATGTGACATTTGGGATTCGCCCAT 2068

RESULT 2

HSA272050

LOCUS HSA272050 3756 bp mRNA linear PRI 01-AUG-2002

DEFINITION Homo sapiens mRNA for transcription initiation factor IA protein (TIF-IA gene).

ACCESSION AJ272050

VERSION AJ272050.1 GI:10046713

KEYWORDS TIF-IA; transcription initiation factor.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Bodem,J., Dobрева,G., Hoffmann-Rohrer,U., Iben,S., Zentgraf,H., Delius,H., Vingron,M. and Grummt,I.

TITLE TIF-IA, the factor mediating growth-dependent control of ribosomal RNA synthesis, is the mammalian homolog of Yeast Rrn3p

JOURNAL EMBO Rep. 1 (2), 171-175 (2000)

MEDLINE 21163499

PUBMED 11265758

REFERENCE 2 (bases 1 to 3756)

AUTHORS Bodem,J.J.

TITLE Direct Submission

JOURNAL Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg, GERMANY

FEATURES

source Location/Qualifiers

1..3756

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

1..3756

/gene="TIF-IA"

23..1978

/gene="TIF-IA"

/function="transcription initiation factor"

/codon\_start=1

/evidence=experimental

/product="transcription initiation factor IA protein"

/protein\_id="CAC07955.1"

/db\_xref="GI:10046714"

/db\_xref="GOA:Q9H4F0"

/db\_xref="SPTREMBL:Q9H4F0"

/translation="MAAPLLHTRLPGDRAASSAVKKLGASRTGISNMRALENDFNFSPPRKTVRFGGTVTEVLKLYKKGENTDFILKNQLLPDIKDDQIINWLLFFRSSIMYLTKDFEQLISIIILRLPWLNRSQTVVEEYLAFLNLVSAQTVFLRPLCSMTASHFVPPRVIIKEGDVDSDDDEDNLPANFDTCRALQIIARYVPSPTWFLMPILVEKFPFVRKSERTLECYVHNLLRISVYPTLRHEILLELIEKLLKLVNASRQGIEDABETATQTCGGTDSGLEFNMDDETHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGKVDNGTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCFSKLGFAEAFLEHLMKKLQDPSNPATIRQAAGNYIGSFARAKFIPLITVKSCLDLVLNWLHIYLNNDSGTKAFCDVALHGPFSACQAVFYXFEVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQNLPLKICLPSVVFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGDSVQICTNPLDTPFFPDPCLVKRSKFIIDPIYQWEDMSAEELQEFKKPMKKDIVEDDDDFLKGEVPQNDTVIGITPSSFDTFRSPSSVSGSPVLYMQPSPL"

3'UTR 1979..3756

/gene="TIF-IA"

/evidence=experimental

ORIGIN

Query Match 97.7%; Score 2020; DB 9; Length 3756;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2020; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 47 GCGGTCGCGTTAGTTCGGCCCAATGGCGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 106

Db 1 GCGGTCGCGTTAGTTCGGCCCAATGGCGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 60

QY 107 TCGGGCCGCTTCGTCCTCTGCAGTTAAGAAAGCTGGGCGCGTCGAGGACTGGGATTTCAA 166

Db 61 TCGGGCCGCTTCGTCCTCTGCAGTTAAGAAAGCTGGGCGCGTCGAGGACTGGGATTTCAA 120

QY 167 TATGCGTGCAATTAGAGAAATGACTTTTCAATTTCTCCCCAAGAAAACTGTTTCGTTTGG 226

Db 121 TATGCGTGCAATTAGAGAAATGACTTTTCAATTTCTCCCCAAGAAAACTGTTTCGTTTGG 180

QY 227 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTT 286

Db 181 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTT 240

QY 287 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCT 346

Db 241 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCT 300

QY 347 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 406

Db 301 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 360

QY 407 ATTAAGATTGCCTTGTTGAATAGAAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCT 466

Db 361 ATTAAGATTGCCTTGTTGAATAGAAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCT 420

QY 467 TGGTAATCTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 526

Db 421 TGGTAATCTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 480

QY 527 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 586

Db 481 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 540

QY 587 TGATGAAGATGATAATCTTCTCGCAATTTTGACACATGTCACAGAGCCTTGCAATAAT 646

Db 541 TGATGAAGATGATAATCTTCTCGCAATTTTGACACATGTCACAGAGCCTTGCAATAAT 600

QY 647 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 706

Db 601 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 660

QY 707 ATTTGTTGAAAAATCAGAGAGAACACTGGAATGTTACGTTTCATACTACTAAGGATTAG 766

Db 661 ATTTGTTGAAAAATCAGAGAGAACACTGGAATGTTACGTTTCATACTACTAAGGATTAG 720



Db	138	TATGCGTGCAATTAGAGAAATGAC	197
QY	227	TGGAACGTGACAGAAAGTCTTGCTG	286
Db	198	TGGAACGTGACAGAAAGTCTTGCTG	257
QY	287	GTTGAAGAACCAAGCTGTTAGATCC	346
Db	258	GTTGAAGAACCAAGCTGTTAGATCC	317
QY	347	AGAATTCGGTTCCTCTATCATGTA	406
Db	318	AGAATTCGGTTCCTCTATCATGTA	377
QY	407	ATTAAGATTGCGCTTGGTTGAATP	466
Db	378	ATTAAGATTGCGCTTGGTTGAATP	437
QY	467	TGGTAATCTTGATPACAGCACAG	526
Db	438	TGGTAATCTTGATPACAGCACAG	497
QY	527	CCATTTTGTGCGTCCCGAGTGAT	586
Db	498	CCATTTTGTGCGTCCCGAGTGAT	557
QY	587	TGATGAAGATGATAAATCTTCCT	646
Db	558	TGATGAAGATGATAAATCTTCCT	617
QY	647	AGCAAGATATGTACCATCGACAC	706
Db	618	AGCAAGATATGTACCATCGACAC	677
QY	707	ATTGTGTCGAAAATCAGAGAAAC	766
Db	678	ATTGTGTCGAAAATCAGAGAAAC	737
QY	767	TGTATATTTTCCAAACCTTGAGG	826
Db	738	TGTATATTTTCCAAACCTTGAGG	797
QY	827	GTTGGATGTGAATGCATCCCGG	886
Db	798	GTTGGATGTGAATGCATCCCGG	857
QY	887	TTGTGGTGGACAGATTCCACGG	946
Db	858	TTGTGGTGGACAGATTCCACGG	917
QY	947	ACATGAACAAAGGCTGGTCCTG	1006
Db	918	ACATGAACAAAGGCTGGTCCTG	977
QY	1007	CCTGGACATCCTGATGTCCTTG	1066
Db	978	CCTGGACATCCTGATGTCCTTG	1037
QY	1067	TAAGGTTGATAACGGCAAAACA	1126
Db	1038	TAAGGTTGATAACGGCAAAACA	1097
QY	1127	ACTCCTGTTGCCACCCCATGCT	1186
Db	1098	ACTCCTGTTGCCACCCCATGCT	1157
QY	1187	TTTCAAATTTGGGATTTCGACAG	1246
Db	1158	TTTCAAATTTGGGATTTCGACAG	1217
QY	1247	AAGTAATCCTGCCATCATCAGGC	1306
Db	1218	AAGTAATCCTGCCATCATCAGGC	1277
QY	1307	AGCTAAATTTTATTCCTCTTAT	1366
Db	1278	AGCTAAATTTTATTCCTCTTAT	1337
QY	1367	GCACATATACCTTAATAACCA	1426
Db	1338	GCACATATACCTTAATAACCA	1397
QY	1427	TGGACCATTTTACTCAGCCTG	1486
Db	1398	TGGACCATTTTACTCAGCCTG	1457
QY	1487	GCTTTTGGCGGAAACCTGAA	1546
Db	1458	GCTTTTGGCGGAAACCTGAA	1517
QY	1547	GATAGTGATGAGCCAGCTAA	1606
Db	1518	GATAGTGATGAGCCAGCTAA	1577
QY	1607	TGCTGCAATCACAAATAAG	1666
Db	1578	TGCTGCAATCACAAATAAG	1637
QY	1667	TCGCCAGATGCTGCCAGTCA	1726
Db	1638	TCGCCAGATGCTGCCAGTCA	1697
QY	1727	AAACCCGCTGGACACACCTT	1786
Db	1698	AAACCCGCTGGACACACCTT	1757
QY	1787	CATTGATCCTATTATTATCAG	1846
Db	1758	CATTGATCCTATTATTATCAG	1817
QY	1847	GAACCCCATGAAAGGACATAG	1906
Db	1818	GAACCCCATGAAAGGACATAG	1877
QY	1907	GCCCCAGATGATACCGTGAT	1966
Db	1878	GCCCCAGATGATACCGTGAT	1937
QY	1967	TCCTTCAAGTAGTGTGGGCT	2026
Db	1938	TCCTTCAAGTAGTGTGGGCT	1997
QY	2027	GCAGAAATTTGTGACTGAGAT	2068
Db	1998	GCAGAAATTTGTGACTGAGAT	2039
RESULT 4			
BC036182			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
BC036182			
Homo sapiens RNA polymerase I transcription factor RRN3, mRNA (CDNA			
clone IMAGE:5298581), partial cds.			
BC036182			
BC036182.1 GI:23243332			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 3709)			
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,			
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,			
Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,			
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,			
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,			
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,			







misc_feature	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/clone_lib="RP11-11K1"	
	/clone_lib="RP11-11 Human Male BAC"	
	1. .1516	
	/note="assembly_fragment"	
misc_feature	1617. .3136	
	/note="assembly_fragment"	
misc_feature	3237. .5148	
	/note="assembly_fragment"	
misc_feature	5249. .7742	
	/note="assembly_fragment"	
misc_feature	7843. .10650	
	/note="assembly_fragment"	
misc_feature	10751. .16194	
	/note="assembly_fragment"	
	clone_end:SP6	
	vector_side:right	
misc_feature	16295. .25751	
	/note="assembly_fragment"	
misc_feature	25852. .37209	
	/note="assembly_fragment"	
misc_feature	37310. .47828	
	/note="assembly_fragment"	
misc_feature	47929. .64239	
	/note="assembly_fragment"	
misc_feature	64340. .77846	
	/note="assembly_fragment"	
misc_feature	77947. .95311	
	/note="assembly_fragment"	
misc_feature	95412. .108930	
	/note="assembly_fragment"	
misc_feature	109031. .135873	
	/note="assembly_fragment"	
ORIGIN		
Query Match 58.6%; Score 1211.6; DB 2; Length 135873;		
Best Local Similarity 80.2%; Pred. No. 0;		
Matches 1617; Conservative 0; Mismatches 334; Indels 65; Gaps 14;		
QY	60	TTGCGCCCAATGGCGGCAACCGCTGCTTCACACGCGTTTGCCGGGAGATGGCGCGCTTCG 119
Db	96731	TTAAGCCCCGGGCAACCGGCTGCTTCTCACATGGTTGCCGGG-GAAGCAGAGGCTTCA 96789
QY	120	TCCTCTGCAGTTAAGAAGCTGGCGCGCTCGAGGACTGGGATTTCAAAATATGCGTGCAATTA 179
Db	96790	GCTTCTGTGGTCAAGAAGCTGGGTATGTCTATGGACCAGTATTTTCAGATATGCTTGCATTC 96849
QY	180	GAGAAATGACTTTTTCAAATTCCTCCCCCAAGAAAAACTGTTCGGTTTGGTGAAGTGTGACA 239
Db	96850	AAGAATGATTTTTCAGCTCTCCCCCAAGAAAAATTTGTTTCATCTTGATGGAACGGGGACA 96909
QY	240	GAACTCTTGCTGAAGTACAAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCCAG 299
Db	96910	GAACTCTTGCGGAAGGACAAAGAGTGAACAAAAATTTAACTTTGAGCTGTTGAAGAGCCAG 96969
QY	300	CTGTTAGATCCAGACATAAAGGATGACCCAGATCATCAACTGGCTGTAGAAATCCGTTCT 359
Db	96970	TTGTGAGATTCAAACATAAAGGATGACCCAGATCTTGAAGTGGCTGTGAAGAAATTTTGCTCT 97029
QY	360	TCTATCATGPACTTTGACAAAAGACTTTGAGCAACTTATCAGTATTTATTAAGATTGCGCT 419
Db	97030	TCTATCATGPACTTTGGCAAAAGACTTTTGAGCAGCTCCTCAGTATTTACATTTGAGATTGCGCT 97089
QY	420	TGGTTGAATPAGAAGTCAACACAGTAGTGGAGAGTATTTGGCTTTTCTTGGTAATCTTGTA 479
Db	97090	TGGTTGAATPAGAATTCAAACAGTGGTGGAGAGTATTTGGCTTCTCTGGTAGTCTTGTA 97149
QY	480	TCAGCACAGACTGTTTTCTCCAGACCGGTGCTCAGCATGATGCTTCCCAATTTGTG--- 536
Db	97150	ACAACAGA---CTCTTCTTAGCCCCAAGTCTCAGCATAAATGCTTCCCAATTTGGCGTAC 97205
QY	537	CCTCCCCGAGTGATCATTAAGGAAGGCGATGTPAGATGTTTCAGATTCTGATGATGAAGAT 596

Db	97206	CCTCCTTGAATGATCATTAAGAAGAGCCATGTAAATGTTTTCAGATTCTGATGATGGAGAT 97265
QY	597	GATAATCTTCCCTGCAAAATTTTGACACATGTACACAGAGCCTTGCAAAATAATAGCAAGATAT 656
Db	97266	GATTATCTTCCCTGCAGATTTTGACACACGTCATACAACCTTACAAAATAACAGCAAGATAT 97325
QY	657	GTACCATCGACACCGTGGTTTCTCATGCCAATACTACCTGGTGGAAAAATTTCCATTGTTTGA 716
Db	97326	GTCCCACTGGCACTGTGGTTTCTTTTGGCCCATACTACCTGGTGGAAACATTTCCATTGTTTC-- 97383
QY	717	AAATCAGAGAGAACACTGGAATGTTACGTTTCATACTACTAAGGATTAGTGTATATTTT 776
Db	97384	-AATCAGAGAGAACACTGGAATGTTTACATTCATACTTACTAAAGATTAGTATATACTTTT 97442
QY	777	CCAACTTTGAGGCATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTG 836
Db	97443	CCAACTTTGAGGAATGAAATTTCTGAAGCTTATTATTAAAAATCTATTCAAAATTTGGATGTG 97502
QY	837	AATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTGGG 896
Db	97503	AATACATTTCTGGCATGATATTGAAGATGCTGAGGAAACAGCAGCTCAAACTTGTGTGGG 97562
QY	897	ACAGATTTCCACGGAGGATTTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAACA 956
Db	97563	ACAGATTTCCATGGAATGATTTGTTCAATATGAACGAAGATGAAGAGAC-----TGTTTCT 97616
QY	957	AAGGCTGGTCTCTGAACGGCTCGACCATGGTGATGCTGCTGTAGCCGAGCGCCTGGACATC 1016
Db	97617	AAGGCTGACTCCAATGGCTCGGTGAGATGGTGCAACCTGTAGCTGAATGCTGGACACC 97676
QY	1017	CTGATGCTTTTGGTTTGTCTACATGAAGGATGTCTGCTATG-----TAGATGGTAAG 1071
Db	97677	CTGCTGCTTTTGTCTTTGTCAGTACTTGAAGTATGTCTGCTATGTAAAAATACATGGTAAG 97736
QY	1072	TTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCC 1131
Db	97737	TTGATAACAAACAAATAAAGGATTTTATA-----ATATATCACAAACTGATAAACATA 97788
QY	1132	TGTTGCCCAACCCATGCTCCTGCCATGTACAGTTTTTTTCATGTTTTTACCTCTGTAGTTTCA 1191
Db	97789	TTTGACAAACCCCTGCTCCTGTACGTACAGTTTTTTTATGTTTTTACCTCTGTGGCTTAA 97848
QY	1192	AATTTGGATTGCGCAGGACATTTTGGAAACATCTCTGG---AAAAATTCAGGACCCAA 1248
Db	97849	AATTTGCACTCTCAGAAAGCATTTTGGAAACGCTCTCTGGGAAAAAAAATGCAGGATCCAA 97908
QY	1249	GTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATATATTTGGAAGCTTTTGGCAAGAG 1308
Db	97909	ATAATCCTGTCTCATCATCAGGTGAGCTGTGGAAATATATTTGGAAGCTCTTTGGCAAAAG 97968
QY	1309	CTAAATTTATTCCTCTTATT--ACTGTAAATCATGCTAGATCTTTTGGTTAACTGSC 1365
Db	97969	CTAAATTCATTCCTCTCATTAACAACATAAAATCATGCTAGGTCTTTTCGGTTAACTGSC 98028
QY	1366	TGCACATATACCTTAATAACAGGATTCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCC 1425
Db	98029	CACACAAATACCTTACTAACCAGGATCTGAGAATGAAGCTTTTGTGATATTGCTCTCC 98088
QY	1426	ATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTTGTTTTAGACACAAGC 1485
Db	98089	ATCAACATTTTACTCAGCCTTCCAAGTGTGTTCCACACCTTTTGTTTTAGACACAAGC 98148
QY	1486	AGCTTTTGAAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGC 1545
Db	98149	AACCTTTAAGCGTAAACCTGAAAGAGGTTTGAGATGTCTTCAGTGCTACATTTTAAAG 98208
QY	1546	GGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTTGCCTGCCCTCAGTGGTTAACTTTT 1605
Db	98209	GGATAGTGACTAGGCAGCTAAATCCTCTGAATATTTGCCTGCCCTCAGTGGTTAACTTAT 98268
QY	1606	TTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACA 1665







Db	20575	AATACATTCTGGCATGATATTGAAGATGCTGAGGAAACAGCAGCTCAAACTTGTGTTGGG	20634
QY	897	ACAGATTCCACGGAGGATTGTTTAATATGATGAAGATGAAGAAACTGAACATGAACA	956
Db	20635	ACAGATTCCATGGAATGATTGTTCAATATGAACGAAGATGAAGAGAC-----TGTTTCT	20688
QY	957	AAGGCTGGTCCCTGAACGGCTCGACCAAGATGGTGCATCCTGTAGCGAGCGCCTGGACATC	1016
Db	20689	AAGGCTGACTCCAAATGGCTCGGTGAGATGGTGACCCCTGTAGCTGAATGCCTGGACACC	20748
QY	1017	CTGATGTCTTTGGTTTTGTCCTFACATGAAGGATGCTGCTATG-----TAGATGGTAAGG	1071
Db	20749	CTGCTGTCTTTGCTTTTGCACTTGAAGTATGCTGCTATGTAATAATACATGGTAAGG	20808
QY	1072	TTGATAACGGCAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTTTGACAAACTCC	1131
Db	20809	TTGATAACAAACAAAATAAAGGATTATA-----ATATATCACAAACTGATAAACATA	20860
QY	1132	TGTTGCCACCCATGCCTCCCTGCCATGTACAGTTTTTTCATGTTTTTACCTCTCTGAGTTTCA	1191
Db	20861	TTTGACAAACCCCTGCCTCCTGTACAGTACAGTTTTTTATGTTTTTACCTCTCTGGCTTAA	20920
QY	1192	AATTGGGATTGCGCAGAGGCAATTTTGTGAACATCTCTGG---AAAAAATGCGAGACCCAA	1248
Db	20921	AATTGCGACTCTCAGAAGCAATTTTGTGAACGCTCTGTGGAAAAAATAATGAGGATCCAA	20980
QY	1249	GTAATCCTGGCATCATCAGGAGGCTGCTGGAATATATATGGAAGCTTTTGGCAAGAG	1308
Db	20981	ATAATCCTGTCAATCATCAGGTGAGCTGCTGGAATTTATTTGGAAGCTCTTTGGCAAAAG	21040
QY	1309	CTAAATTTATTCCTCTTATT---ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGC	1365
Db	21041	CTAAATTCATTCCTCTCTCATTAACAACATAAAATCATGCCTAGTCTTTTCGTTAACTGGC	21100
QY	1366	TGCATATATACCTTAATAACCAAGGATTCGGGAACAAAGGCATCTTCGATGTTGCTCTCC	1425
Db	21101	CACACAAATACCTTACTAACCAGGATCTGAGAATGAAGGCTTTTGTGATATGCTCTCC	21160
QY	1426	ATGGACCATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGC	1485
Db	21161	ATCAAAACATTTTACTCAGCCTTCCAAGCTGTGTTCCACACCTTTGTTTTTAGACACAAGC	21220
QY	1486	AGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGC	1545
Db	21221	AACTTTTAAAGCGTAAACCTGAAAGAAGGTTTGAGATGTCTTCAGTGCCTACATTTTAAAG	21280
QY	1546	GGATAGTGATGAGCCAGCTAAATCCCTGAAGATTTTGGCTGCCCTCAGTGGTTAACTTTT	1605
Db	21281	GGATAGTGACTAGGCAGCTAAATCCTCTGAATATTTTGGCTGCCCTCAGTGGTTAACTTAT	21340
QY	1606	TTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGTACACCATCATTTGAGAGGAACA	1665
Db	21341	TTGCTGTGTACAGAAATAAATATCAGCTAG--TCTGTACACTATCATTTGAGAGGAACA	21397
QY	1666	ATCGCCAGATGCTGCCAGTCAATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCA	1725
Db	21398	ACGGCCAGA-----AATACAGCTGGAGGAGACTCGGTGTGAACGTGCA	21440
QY	1726	CAAACCCGCTGGACACCTTCTTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAAT	1785
Db	21441	CCAACCATTTGGACGCTTCTTCCCTTTGATCCCTTGTGCGCTTAAACAGTCAAGGAAAG	21500
QY	1786	TCATTGATCCTATTATTACAGGTATGGGAAGACATGATGCTGAAGACTACAGGAGTTCA	1845
Db	21501	TCATTGATTCTATTATTACAGATATGGGAAGACATAAGTGTGGAAGAGCTTCAGGAGTTTA	21560
QY	1846	AGAAACCCATGAARAAGGACATAGTGGAAAGATGAAGATGATGACTTTCTGAAAGCGCAAG	1905
Db	21561	AGAAACCTATGAARAAGGAGATGGAAAAAGATGGAGATGAT---CTTTTGAAGGTGAAG	21617
QY	1906	TGCCCCAGAATGATACCGTGATTGGGATCACACCAGCTCCTTTGACACG-----CATTT	1960
Db	21618	TTCCCCAAATGATACCATGATCGGGATTATACCGAGCCACTTTTGATGCGATGCACATTT	21677

QY	1961	CCGAAGTCCTTCAAGTAGTGTGGGCTCCCCACCCGCTGTTGTATCATG-CAACCCAGTCCCC	2019
Db	21678	CCAAAGTCCTTCAAGTAGTGTGGGCTCCCCACCTGTGTACACTGCCAGACCAGTTCCT	21737
QY	2020	TCTGACGGCAGAAATTTTGTGACTGAGATGTGACATT	2055
Db	21738	TTCATCTCGAGGATTTGTGATGAGATGCAGCATT	21773

RESULT 7			
LOCUS	BC034110		
DEFINITION	Mus musculus expressed sequence AL0233001, mRNA	linear	ROD 16-APR-2003
ACCESSION	BC034110	IMAGE:4504751, partial cds.	
VERSION	BC034110.1	GI:21706612	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		

REFERENCE	1 (bases 1 to 3166)		
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., Kornman,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903	(2002)
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 3166)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		

REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk
	Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a>
	Tissue Procurement: The Cepko Laboratory
	cDNA Library Preparation: Life Technologies, Inc.
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
	Web site: <a href="http://www-shgc.stanford.edu">http://www-shgc.stanford.edu</a>
	Contact: (Dickson, Mark) <a href="mailto:mcd@paxil.stanford.edu">mcd@paxil.stanford.edu</a>
	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 44 Row: 9 Column: 5  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Location/Qualifiers

```
source
1. .3166
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:504751"
/tissue_type="Eye, retina, mouse strain C57Bl\6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
<1. .3166
/gene="AL023001"
/db_xref="LocusID:106298"
/db_xref="MGI:2146525"
<1. .1544
/gene="AL023001"
/codon_start=3
/product="AL023001 protein"
/protein_id="AAH34110.1"
/db_xref="GI:21706613"
/db_xref="LocusID:106298"
/db_xref="MGI:2146525"
/translation="PCLSMIAshFVPPRIVIVKEGGIDVSDSDDEDDNLPAIFDTCRA
LQIITRVYVPTPWFLLPIIVKFPFVRKSERTLECYVHNLRLISLYFPTLRREILELV
IEKLLDVSVSRQDIEDAEKAAQTCGGTDTTEGLFNMDDEDEDTPKKADQEQPNQ
MAHPTAERLDVLLCLLSYIEDVCRVHGKIDNNKTKOLYRDLISIFDKLLLPHTASCH
VQFMFFLCSFKLGFAEFLHLWKLLQDPNNPAIIRQAAANYIGSFLARAKFIPLIT
VKTCLDLVNLHMYLTNODSGTKAFCDVALGHPFYSACQAVFYTVVFRHKQLLSGNL
KQGLQYLSLNFPERIVLSQNLPLKICLPQVNVFFFAAIITNKYQLVFCYTIMERNRQML
PQVIRSTAGDSVQTCNPLDTPFPDFPCVLKRSKKFIDPIYQIWDGSAEELQEFKKS
TKKEVVEDEDDDLKGEVPSQSDTVTGLTSSFDTHFQSPSSSVGSPVLIYIPGQSPLL
TRIYD"
```

ORIGIN

Query Match 57.1%; Score 1179.8; DB 10; Length 3166;  
Best Local Similarity 85.3%; Pred. No. 0;  
Matches 1327; Conservative 0; Mismatches 227; Indels 1; Gaps 1;

```
Qy 502 GACCGTGCTCAGCATGATTGCTTCCCATTTTGTGCGCTCCCGAGTGATCATTAAAGGAAG 561
Db 1 GACCATGTCTCAGCATGATTGCGTCTCATTTTGTACCTCCCGAGTAATTGTCAAGGAAG 60

Qy 562 GCGATGATAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCTCGAAAATTTTGACA 621
Db 61 GTGGCATAGATGTTTCAGATTCTGATGACGAAGATGACAATCTTCTCGCAATTTTGTACA 120

Qy 622 CATGTCAAGAGCCCTTGAAAATAATAGCAAGATATGTACCATCGACACCGTGGTTTCTCA 681
Db 121 CATGTCAAGAGCCCTTGAAAATAATACAAAGATATGTCCCATCGACACCATGGTTTCTAA 180

Qy 682 TGCCAATACTGGTGGAATAATTTCCATTTGTTGCGAAATACAGAGAGAACACTGGAATGTT 741
Db 181 TGCCAATACTGGTGGAATAATTTCCCGTTTGTGAGGAAGTCCGAGAGAACATTGGAATGTT 240

Qy 742 ACGTTCATAAATTACTAAGGATAGTGATATATTTTCCAACTTGAGGCATGAAATTTCTGG 801
Db 241 ATGTTCAATACTTATTAGGATAAGTTTATATTTCCCAACTTTGAGGCGTGAAATTTCTGG 300

Qy 802 AGCTTATTATTGAAAAAATACTCAAGTTGATGTAATGATGATCCCGGCGAGGATATTGAAG 861
Db 301 AGCTTGTTATTGAAAAGCTACTCAAGTTAGATGTGATGTATCGCGGCGAGGATATTGAAG 360

Qy 862 ATGCTGAAGAAAACAGCAACTCAAACTTGTTGGTGACAGATCCACGGAAGGATTGTTTA 921
Db 361 ATGCTGAAGAAAAGCAGCTCAGACTTGTTGGTGGGACAGATACCACAGAGGACTGTTTA 420

Qy 922 ATATGGATGAAGTGAAGAAACTGAACATGAACAAAGCGTGGTCTCTGAAACGGCTCGACC 981
Db 421 ATATGGATGAAGATGAGGACACTGACCTGAGAAAGAAAGCTGACACGAGGAGCCTTAACC 480

Qy 982 AGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCCTTTGGTTTGTCTCTACA 1041
Db 481 AGATGGCTCATCCCAACCGCAGAGCGCCTGGACGTCCTGCTGTGCTTGTCTGTCTCTACA 540
```

```
Qy 1042 TGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATC 1101
Db 541 TTGAGGATGTCTGCCGTGTGCACGGTAAATTTGATAACAATAAAACAAAGGATTTATACC 600

Qy 1102 GCGACCTGATAAAACATCTTTGACAAACTCTCTGTTGCCCAACCCATGCCCTCCTGCCATGTAC 1161
Db 601 GTGATCTGATATCCATCTTTGACAAACTTCTGTTGCCCAACACATGCTCCTCCTGCCATGTAC 660

Qy 1162 AGTTTTTCATGTTTACCTCTGTAGTTTCAAAATTTGGGATTCGCAGAGGCAATTTTGGAAAC 1221
Db 661 AGTTCTTCATGTTTTCCTCTGCAGCTTCAAGTTGGGATTTGCAGAAAGCAATTTTGGAAAC 720

Qy 1222 ATCTCTGGAAAAAATTCAGGACCCCAAGTAATCCTGCCCATCATCAGGCAGGCTGCTGGAA 1281
Db 721 ATCTCTGGAAAAAGTTGCAGGATCCAAATAACCCGCCCATCATCAGGCAGGCTGCTGCAA 780

Qy 1282 ATTATATTGGAAAGCTTTTGGCAAGAGCTAAATTTATTCTCTTATTACTGTAAAAATCAT 1341
Db 781 ATTATATTGGGAGCTTTTGGCCAGAGCTAAATTTATTCTCTTATCACTGTGAAGACAT 840

Qy 1342 GCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAAGGATTCGGGAACAA 1401
Db 841 GCCTGGATCTCTTGGTTAACTGGCTGCACATGTACCTTACTAACCAAGGATTCGGGAACAA 900

Qy 1402 AGGATTTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCT 1461
Db 901 AGGCTTTTGTGACGTTGCACTCCATGGACCATTTTATTTCAGCCTGCCAAGCTGTGTTCT 960

Qy 1462 ACACCTTTGTTTGTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAGAAGAGTTTGCAGT 1521
Db 961 ACCTGTTGTTTGTAGACACAAGCAGCTTTTGAGTGGAAACTTGAAGCAAGGTTCTACAGT 1020

Qy 1522 ATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCTGGAAGATTT 1581
Db 1021 ATCTTCAGAGTCTAAATTTTGAGCGCATTTGTGTTGAGCCAGCTGAACCCACTGAAGATCT 1080

Qy 1582 GCCTGCCCTCAGTGGTTAACTTTTGTGCTGCAATCAAAAATAAGTACCAAGCTGCTTCT 1641
Db 1081 GCCTGCCGCAAGTGGTTAAATTTCTTGTGCTATCAAAAATAAATACCAGCTGGTGTCT 1140

Qy 1642 GCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCAATTAGGAGTACCGCTG 1701
Db 1141 GCTACACCATCATTTGAGAGGAACAAGTCCAGATGCTCCCGTTATCCGCAGCACCGCTG 1200

Qy 1702 GAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTCCCTTTTGATCCCT 1761
Db 1201 GTGGAGACTCCGTGCAGACCTGCACCAACCCACTGGACACTTCTTCCCTTTTGACCCCTT 1260

Qy 1762 GTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTATGGGAAGACATGA 1821
Db 1261 GTGTGCTTAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGATTGGGAAGATGGGA 1320

Qy 1822 GTGCTGAAGAGCTACAGGAGTTCAAGAAACCCCATGAAAAGGACATAGTGAAGATGAAG 1881
Db 1321 GTGCTGAAGAGCTTCAGGAGTTTAAGAAATCTACTAAAAGGAGGTAGTGGAGGATGAAG 1380

Qy 1882 ATGATGACTTTCTGAAAGGGAAGTGCCTTCAAGTAGTGTGGGCTCCCAACCCGTTGTGT 1941
Db 1381 ATGATGACTTTTGAAGGCGAGGTGCCCCAGAGTGCACACAGTACTGGCCTTACTCCGA 1440

Qy 1942 GCTCCTTTGACACGCAATTCGGAAGTCCCTTCAAGTAGTGTGGGCTCCCAACCCGTTGTGT 2001
Db 1441 GCTCCTTTGATACCCACTTCCAAAGTCCCTCCAGTAGTGTGGGCTCCCTCCTGTGTGTGT 1500

Qy 2002 ACAT-GCAAACCCAGTCCCTCTGACGGCAGAAAATTTGTGACTGAGATGTGACATT 2055
Db 1501 ATATACCAGGCCAGTCTCCACTCCTCACAAGGATCTATGATGAGTGGCGCAATT 1555
```

RESULT 8

AX427479

LOCUS

DEFINITION

AX427479 1423 bp DNA linear PAT 20-JUN-2002  
Sequence 52 from Patent WO01233558.





JOURNAL Patent: EP 1293569-A 544 19-MAR-2003;  
Helix Research Institute (Jp) ; Research Association for  
Biotechnology (Jp)  
FEATURES Location/Qualifiers  
source 1. .2493  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 27.4%; Score 567; DB 6; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 6.6e-145;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1196 GGGATTGCGAGAGGCATTTTGGAAACATCTCTGGAAAAAATTGACGACCCCAAGTAATCC 1255  
Db 711 GGGATTGCGAGAGGCATTTTGGAAACATCTTTGGAAAAAATTGACGATCCAAGTAATCC 770  
QY 1256 TGCCATCATCAGGCGCTGCTGGAAATATATTGGAAGCTTTTGGCAAGAGCTAAATT 1315  
Db 771 TGCCATCATCAGGCGCTGCTGGAAATATATTGGAAGCTTTTGGCAAGAGCTAAATT 830  
QY 1316 TATTCCTCTTATTACTGTAAAAATCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATA 1375  
Db 831 TATTCCTCTTATTACTGTAAAAACCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATA 890  
QY 1376 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTCCTGGCATGTTGCTCTCCATGGACCAAT 1435  
Db 891 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTCCTGAATGTTGCTCTCCATGGACCAAT 950  
QY 1436 TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTGTTTGTAGACACAGCAGCTTTTGAG 1495  
Db 951 TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTGTTTGTAGACACAGCAGCTTTTGAG 1010  
QY 1496 CGGAAACTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 1555  
Db 1011 CGGAAACTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 1070  
QY 1556 GAGCCAGCTAAATCCCTGAAAGATTGCGCTGCCCTCAGTGGTTAACTTTTGTGCAAT 1615  
Db 1071 GAGCCAGCTAAATCCCTGAAAGATTGCGCTGCCCTCAGTGGTTAACTTTTGTGCAAT 1130  
QY 1616 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATFTGAGAGGAACAATCGCCAGAT 1675  
Db 1131 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATFTGAGAGGAACAATCGCCAGAT 1190  
QY 1676 GCTGCCAGTCAATTAGGATACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCT 1735  
Db 1191 GCTGCCAGTCAATTAGGATACCACTGGAGAGACTCAGTGCAGACCTGCACAAACCCCACT 1250  
QY 1736 GGACACCTTCTTCCCTTTGATCCCTGTGCTGAAGAGGTCA 1778  
Db 1251 GGACACCTTCTGTCCTCCCTTTGATCCCTGTGCTGAAGAGGTGA 1293

RESULT 10  
AK055742  
LOCUS AK055742 2493 bp mRNA linear PRI 01-AUG-2002  
DEFINITION Homo sapiens cDNA FLJ31180 fis, clone KIDNE2000266.  
ACCESSION AK055742  
VERSION AK055742.1 GI:16550546  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,  
Ono,Y., Hotuta,T., Hirakawa,S., Murakawa,K., Takiguchi,S.,  
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,  
Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,  
Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,  
Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,

Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2493)  
Isogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.  
FEATURES Location/Qualifiers  
source 1. .2493  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="KIDNE2000266"  
/tissue type="kidney"  
/clone\_lib="KIDNE2"  
/note="cloning vector: pME18SFL3"  
ORIGIN  
Query Match 27.4%; Score 567; DB 9; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 6.6e-145;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1196 GGGATTGCGAGAGGCATTTTGGAAACATCTCTGGAAAAAATTGACGACCCCAAGTAATCC 1255  
Db 711 GGGATTGCGAGAGGCATTTTGGAAACATCTTTGGAAAAAATTGACGATCCAAGTAATCC 770  
QY 1256 TGCCATCATCAGGCGCTGCTGGAAATATATTGGAAGCTTTTGGCAAGAGCTAAATT 1315  
Db 771 TGCCATCATCAGGCGCTGCTGGAAATATATTGGAAGCTTTTGGCAAGAGCTAAATT 830  
QY 1316 TATTCCTCTTATTACTGTAAAAATCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATA 1375  
Db 831 TATTCCTCTTATTACTGTAAAAACCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATA 890  
QY 1376 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTCCTGGCATGTTGCTCTCCATGGACCAAT 1435  
Db 891 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTCCTGAATGTTGCTCTCCATGGACCAAT 950  
QY 1436 TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTGTTTGTAGACACAGCAGCTTTTGAG 1495  
Db 951 TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTGTTTGTAGACACAGCAGCTTTTGAG 1010  
QY 1496 CGGAAACTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 1555  
Db 1011 CGGAAACTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 1070  
QY 1556 GAGCCAGCTAAATCCCTGAAAGATTGCGCTGCCCTCAGTGGTTAACTTTTGTGCAAT 1615  
Db 1071 GAGCCAGCTAAATCCCTGAAAGATTGCGCTGCCCTCAGTGGTTAACTTTTGTGCAAT 1130  
QY 1616 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATFTGAGAGGAACAATCGCCAGAT 1675  
Db 1131 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATFTGAGAGGAACAATCGCCAGAT 1190  
QY 1676 GCTGCCAGTCAATTAGGATACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCT 1735  
Db 1191 GCTGCCAGTCAATTAGGATACCACTGGAGAGACTCAGTGCAGACCTGCACAAACCCCACT 1250  
QY 1736 GGACACCTTCTTCCCTTTGATCCCTGTGCTGAAGAGGTCA 1778  
Db 1251 GGACACCTTCTGTCCTCCCTTTGATCCCTGTGCTGAAGAGGTGA 1293



RESULT 11  
BC009198  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1168)  
Klausner,R.D., Collins,F.S., Wagner,L.H., Derge,J.G.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 1168)  
Strausberg,R.  
Direct Submission  
Submitted (06-JUN-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Aug 19, 2003 this sequence version replaced gi:14327947.  
Contact: MGC help desk  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ruben Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>  
contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 24 Row: k Column: 14.

FEATURES  
source

Location/Qualifiers  
1. .1168  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3678732"  
/tissue\_type="Lymph, Burkitt lymphoma"  
/clone\_lib="NIH MGC 8"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
<1. .627  
/codon\_start=1  
/product="Unknown (protein for IMAGE:3678732)"

CDS

/protein\_id="AAH09198.2"  
/db\_xref="GI:33872205"  
/translation="RSGRVSSAQWRHRCFTRCLSGDAASSAVKTLGASRTGTYLNE  
DKNLGRARLLTPVIPALWEAEVGGSPVRSRPAWPTCETSSLEKIQKLARRGGRLSN  
MRALENDFFNSPPRKTVRFGGTVTEVLLKYKKGENTDFELLKNQLDPPDKDDQIINW  
LLEFRSSVMYLTAKDFEQLISIIILRLPWLNRSQTVVEYLAFGLNLVSA"

ORIGIN

Query Match 24.5%; Score 506.4; DB 9; Length 1168;  
Best Local Similarity 88.5%; Pred. No. 3.2e-128;  
Matches 549; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 157 GGATTTCAAATATGCGTGCAATTAGAGAAATGACTTTTTCAAATTCTCCCCAAGAAAAAAGT 216  
Db |||||  
QY 296 GGATTTCAAATATGCGTGCAATTAGAGAAATGACTTTTTCAAATTCTCCCCAAGAAAAAAGT 355  
Db |||||  
QY 217 TTCGGTTTGGTGGAACTGTGACAGAACTCTTGCTGAAGTACAAAAAGGGTGAAAAAAGT 276  
Db |||||  
QY 356 TTCGGTTTGGTGGAACTGTGACAGAACTCTTGCTGAAGTACAAAAAGGGTGAAAAAAGT 415  
Db |||||  
QY 277 ACTTTGAGTTGTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCA 336  
Db |||||  
QY 416 ACTTTGAGTTGTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCA 475  
Db |||||  
QY 337 ACTGGTCTAGAAATCCGTTCTTCTATCATGTACTTGACAAAAAGACTTTGAGCAACTTA 396  
Db |||||  
QY 476 ACTGGTCTAGAAATCCGTTCTTCTGTCATGTACTTGACAAAAAGACTTTGAGCAACTTA 535  
Db |||||  
QY 397 TCAGTATTATATTAAAGATTGCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATT 456  
Db |||||  
QY 536 TCAGTATTATATTGAGATTGCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATT 595  
Db |||||  
QY 457 TGGCTTTTCTGGTAATCTTGATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCA 516  
Db |||||  
QY 596 TGGCTTTTCTGGTAATCTTGATCAGCATAGACTGTTTTCCTCAGACCGTGTCTCAGCA 655  
Db |||||  
QY 517 TGATTGCTTCCCATTTTGTGCCTCCCAGAGTGATCATTAAGGAAGGCGATGTAGATGTTT 576  
Db |||||  
QY 656 TGATTGCTTCCCATTTTGTGCCTCCCAGAGTGATCATTAAGGAAGGCGATGTAGATGTTT 715  
Db |||||  
QY 577 CAGATTCTGATGATGAAGATGATAATCTTCTCCTGCAAAATTTGACACATGTCACAGAGCCT 636  
Db |||||  
QY 716 CAGATTCTGATGATGAAGATGATAATCTTCTCCTGCAAAATTTGACACATGTCACAGAGCCT 775  
Db |||||  
QY 637 TGCAATAAATAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGG 696  
Db |||||  
QY 776 TGCAATAAATAGCAAGATATGTACCATCATGAGTATATCTTTCCTTATTTTGAATGTTTA 835  
Db |||||  
QY 697 AAAAAATTCATTGTTGAAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAACTTAC 756  
Db |||||  
QY 836 ATTCTCAAGAAAAATTCATAATCAATTAGTAAAAATTTATAAAATGTTTAAATAGTATTAAGCT 895  
Db |||||  
QY 757 TAAGGATTAGTGATATTTT 776  
Db |||||  
QY 896 TGAGTCTTACATTGCATTTT 915  
Db |||||

RESULT 12  
BC055781

LOCUS

DEFINITION

BC055781

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2243)

Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

BC055781 2243 bp mRNA linear ROD 04-NOV-2003  
Mus musculus expressed sequence AL023001, mRNA (cdna clone  
IMAGE:6403754), partial cds.

BC055781

BC055781.1 GI:33585913

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2243)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2243)

Strausberg,R.

Direct Submission

Submitted (01-AUG-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 125 Row: c Column: 18

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source

Location/Qualifiers

1..2243

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:6403754"

/tissue\_type="Brain, mouse 12.5 dpc"

/clone\_lib="NIH BMAP\_FD0"

/lab\_host="DH10B"

/note="Vector: pYX-ASC"

<1..2243

/gene="AL023001"

/note="synonym: Rrn3"

/db\_xref="LocusID:106298"

/db\_xref="MGI:2146525"

<1..647

/gene="AL023001"

/codon\_start=3

/product="AL023001 protein"

/protein\_id="AAH55781.1"

/db\_xref="GI:33585914"

/db\_xref="LocusID:106298"

/db\_xref="MGI:2146525"

/translation="KAFCDVALHGPFYSACQAVFYTVVFRHKLLSGNLKQGLQYLQS LNFRIVLSQLNPLKICLPQVNVFAAITNKYQLVFCYTIMERNRQMLPVIRSTAGG DSVQTCTNPLDTFFPDPVCLKRSKKFIDPIYQIWEDGSAEELQEFKKSTKKQVVEDE

gene

CDS

misc\_feature

DDDFLKGEVPQSDTVTGLTPSSFDTHFQSPSSVSGSPVLYIPGQSPLLTRIYD" 27..434

/gene="AL023001"

/note="Region: RNA polymerase I specific transcription initiation factor RRN3. This family consists of several eukaryotic proteins which are homologous to the yeast RRN3 protein. RRN3 is one of the RRN genes specifically required for the transcription of rDNA by RNA polymerase I (Pol I) in Saccharomyces cerevisiae"

/db\_xref="CDD:pfam05327"

ORIGIN

Query Match 23.1%; Score 478; DB 10; Length 2243;

Best Local Similarity 83.9%; Pred. No. 2.5e-120;

Matches 552; Conservative 0; Mismatches 105; Indels 1; Gaps 1;

QY 1399 CAAAGGCATTCTGCGATGTTGCTCTCCATGGACCAATTTACTCAGCCTGCCAAGCTGTG 1458

Db 1 CAAAGGCTTTTGTGACGTTGCACTCCATGGACCAATTTATTTCAGCCTGCCAAGCTGTG 60

QY 1459 TCTACACCTTTGTTTTAGACACACAGCAGCTTTTGAGCGGAAACCTGAAAGAGGTTGC 1518

Db 61 TCTACACTGTTGTTTTAGACACACAGCAGCTTTTGAGTGGAAACTTGAAGCAAGGTCTAC 120

QY 1519 AGTATCTTCAGAGTCTGAATTTTGGAGGGGATAGTGATGAGCCAGCTAAATCCCTGAAGA 1578

Db 121 AGTATCTTCAGAGTCTAAATTTTGGAGGCAATTTGTTGAGCCAGCTGAACCCACTGAAGA 180

QY 1579 TTTGCCTGCCCTCAGTGGTTAACTTTTGTGCTGCATCACAAATAAGTACCAGCTCGTCT 1638

Db 181 TCTGCTGCCGCAAGTGTTAAATTTTGTGCTGCTATCACAAATAATACCAGCTGGTGT 240

QY 1639 TCTGTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTATTAGGAGTACCG 1698

Db 241 TCTGTACACCATCATGGAGAGGAACAGTCGCCAGATGCTCCCCGTATCCGCAGCACCG 300

QY 1699 CTGGAGGAGACTCAGTGCAGATCTGCACAAAACCCGCTGGACACCTTCTTCCCTTTGATC 1758

Db 301 CTGGTGGAGACTCCGTCAGACCTGCACCAACCCACTGGACACTTTCTTCCCTTTGACC 360

QY 1759 CCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTATGGGAAGACA 1818

Db 361 CTTGTGTGCTTAAGAGGTCAAAGAAAGTTCATTGATCCTATTTATCAGATTTGGGAAGATG 420

QY 1819 TGAGTGTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAGGACATAGTGAAGATG 1878

Db 421 GGAGTGTGAAGAGCTTCAGGAGTTTAAAGAAA TCTACTAAAAGACAGGTAGTGGAGGATG 480

QY 1879 AAGATGATGACTTTCTGAAAGCGGAGTCCCCAGATGATACCGTATGGGATCACAC 1938

Db 481 AAGATGATGACTTTTGAAGAGCGAGGTGCCCCAGAGTGACACAGTGACTGGCCTTACTC 540

QY 1939 CAAGCTCCTTTGACACGCATTTCCGAAGTCCCTTCAAGTAGTGTGGGCTCCCCACCCGTGT 1998

Db 541 CGAGCTCCTTTGATACCCACTTCCAAAGTCCTTCCAGTAGTGTGGGCTCCCCCTCCTGTGC 600

QY 1999 TGTACAT-GCAACCCAGTCCCCCTCTGACGCGCAGAAATTTGTGACTGAGATGTGACATT 2055

Db 601 TGTATATACACAGGCCAGTCTCCACTCCTCAAGGATCTATGATTGAGCTGCGCAATT 658

RESULT 13

AX274907

LOCUS AX274907 1461 bp DNA linear PAT 29-OCT-2001

DEFINITION Sequence 172 from Patent WO0172777.

ACCESSION AX274907

VERSION AX274907.1 GI:16547539

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1



QY 1428 GGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTTGTTTGTAGACACAGCAG 1487  
|||||  
Db 1932 GGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTTGTTTGTAGACACAGCAG 1991  
|||||  
QY 1488 CTTTGTAGCGGAAACCTGAAAGAGCTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGG 1547  
|||||  
Db 1992 CTTTGTAGCGGAAACCTGAAAGAGCTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGG 2051  
|||||  
QY 1548 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGGCTGCCCTCAGTGGTTAACTTTT 1607  
|||||  
Db 2052 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGGCTGCCCTCAGTGGTTAACTTTT 2111  
|||||  
QY 1608 GCTGCAATCACAAATAAGTA 1627  
|||||  
Db 2112 GCTGCAATCACAAAGATGAA 2131  
|||||

RESULT 15  
AK126166 AK126166 3645 bp mRNA linear PRI 09-SEP-2003  
LOCUS Homo sapiens cDNA FLJ44178 fis, clone THYMU2037233, highly similar  
DEFINITION to Homo sapiens RNA polymerase I transcription factor RRN3 (RRN3).  
AK126166  
ACCESSION AK126166.1 GI:34532564  
VERSION  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1

REFERENCE  
AUTHORS Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,  
Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,  
Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,  
Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,  
Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,  
Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,  
Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.  
and Isogai,T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3645)  
AUTHORS Isogai,T. and Yamamoto,J.  
TITLE Direct Submission  
JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
FEATURES  
source Location/Qualifiers  
1 .3645  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="THYMU2037233"  
/tissue\_type="thymus"  
/clone\_lib="THYMU2"  
/note="cloning vector: pME18SFL3"

ORIGIN  
Query Match 20.3%; Score 420.8; DB 9; Length 3645;  
Best Local Similarity 97.3%; Pred. No. 1.8e-104;  
Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 1188 TTCAAATGGGATTCGAGAGGCAATTTTGGAAACATCTCTGGAAAAAATTCAGGACCCA 1247  
Db 2771 TTAAATATGGGATTCGAGAGGCAATTTTGGAAACATCTTTGGAAAAAATTCAGGATCCA 2830  
|||||

QY 1248 AGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGA 1307  
|||||  
Db 2831 AGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGA 2890  
|||||  
QY 1308 GCTAAAATTTATTCTCTTATTACTGTAAAAATCATGCCTAGATCTTTTGGTTAACTGGCTG 1367  
|||||  
Db 2891 GCTAAAATTTATTCTCTTATTACTGTAAAAACCATGCCTAGATCTTTTGGTTAACTGGCTG 2950  
|||||  
QY 1368 CACATATACCTTAATAACCCAGGATTCGGGAACAAAGGCATTTCTGGATGTTGCTCTCCAT 1427  
|||||  
Db 2951 CACATATACCTTAATAACCCAGGATTCGGGAACAAAGGCATTTCTGGATGTTGCTCTCCAT 3010  
|||||  
QY 1428 GGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTTGTATAGACACAAGCAG 1487  
|||||  
Db 3011 GGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTTGTATAGACACAAGCAG 3070  
|||||  
QY 1488 CTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGG 1547  
|||||  
Db 3071 CTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGG 3130  
|||||  
QY 1548 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGGCTGCCCTCAGTGGTTAACTTTT 1607  
|||||  
Db 3131 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGGCTGCCCTCAGTGGTTAACTTTT 3190  
|||||  
QY 1608 GCTGCAATCACAAATAAGTA 1627  
|||||  
Db 3191 GCTGCAATCACAAAGATGAA 3210  
|||||

Search completed: July 18, 2004, 04:52:49  
Job time : 8122 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 01:13:34 ; Search time 816 Seconds  
(without alignments)  
10766.267 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagaggctgtggctggaag.....tgacatttggtcccat 2068

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2068	100.0	2068	6	ABV72159	Abv72159 Nucleotid
2	2018.4	97.6	2040	3	AAA98384	Aaa98384 Human RNA
3	1457.6	70.5	1770	6	ABS78724	Abs78724 DNA encod
4	1253.6	60.6	1582	9	ADD29697	Adg29697 Human tum
5	1228.4	59.4	1418	3	AAC76937	Aac76937 Human ORF
6	1010.4	48.9	1498	4	AAS44584	Aas44584 Human ful
7	894	43.2	1423	4	AAS01562	Aas01562 Human sec
8	639.4	30.9	2410	5	AAS92255	Aas92255 DNA encod
9	568.6	27.5	776	5	AAS92253	Aas92253 DNA encod
10	567	27.4	2493	6	ABS64829	Abs64829 Human tum
11	567	27.4	2493	7	ADA52976	Ada52976 Human cod
12	550.8	26.6	1260	5	AAS87643	Aas87643 DNA encod
13	477.8	23.1	655	5	AAS80955	Aas80955 DNA encod
14	466.2	22.5	3169	5	AAS87641	Aas87641 DNA encod
15	420.8	20.3	1461	5	ABA83038	Abas83038 Human tra
16	420	20.3	437	2	AAX51663	Aax51663 Human tra
17	419.2	20.3	1989	6	ABK47950	Abk47950 Human tra
18	398	19.2	560	5	AAS91991	Aas91991 DNA encod
19	390.6	18.9	3175	5	AAS92254	Aas92254 DNA encod
20	387	18.7	585	5	AAS92248	Aas92248 DNA encod
21	315	15.2	1017	5	AAS87636	Aas87636 DNA encod
22	292	14.1	341	2	AAV90248	Aav90248 EST clone
23	286.6	13.9	498	5	AAS92247	Aas92247 DNA encod

24	282.2	13.6	396	5	AAS91986	Aas91986 DNA encod
25	282.2	13.6	396	5	AAS80954	Aas80954 DNA encod
26	282.2	13.6	396	5	AAS92246	Aas92246 DNA encod
27	282.2	13.6	396	5	AAS87635	Aas87635 DNA encod
28	279	13.5	2991	5	AAS91992	Aas91992 DNA encod
29	279	13.5	2991	5	AAS73660	Aas73660 DNA encod
30	247.4	12.0	349	3	AAC77533	Aac77533 Human ORF
31	233.4	11.3	667	5	AAS92251	Aas92251 DNA encod
32	230	11.1	2103	3	AAC77798	Aac77798 Human can
33	208	10.1	34269	4	AAK68677	Aak68677 Human imm
34	208	10.1	34269	4	AAK85168	Aak85168 Human imm
35	203.6	9.8	213	3	AAC19091	Aac19091 Human sec
36	192	9.3	17803	4	AAK68676	Aak68676 Human imm
37	188.8	9.1	9259	4	AAK68674	Aak68674 Human imm
38	188.8	9.1	23815	4	AAK85169	Aak85169 Human imm
39	188.8	9.1	23815	4	AAK68678	Aak68678 Human imm
40	163.4	7.9	627	5	AAS87640	Aas87640 DNA encod
C 41	156	7.5	557	6	ABV87866	Abv87866 Human col
42	144.2	7.0	3410	4	AAK85167	Aak85167 Human imm
C 43	139.4	6.7	1326	5	AAS91907	Aas91907 DNA encod
44	139.4	6.7	3042	5	AAS70319	Aas70319 DNA encod
C 45	138.6	6.7	807	5	AAS92252	Aas92252 DNA encod

ALIGNMENTS

RESULT 1  
ABV72159  
ID ABV72159 standard; DNA; 2068 BP.  
XX  
AC ABV72159;  
XX  
DT 05-DEC-2002 (first entry)  
XX  
DE Nucleotide sequence of human RRN3.  
XX  
KW Human; RRN3; RNA polymerase I transcription factor;  
KW hypoproliferative disease; hyperproliferative disease; cancer;  
KW malignancy; hyperplasia; metaplasia; dysplasia; benign tumour;  
KW hypoproliferative disorder; benign dysproliferative disorder;  
KW autoimmune disease; cardiac disease; degenerative disorder;  
KW growth deficiency; hypoproliferative disorder; physical trauma; lesion;  
KW wound; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 69..2024  
/\*tag= a  
/product= "RRN3"  
US2002090706-A1.  
11-JUL-2002.  
16-AUG-2001; 2001US-00932678.  
16-AUG-2000; 2000US-0225893P.  
(REED/) REEDER R H.  
(MOOR/) MOOREFIELD B.  
(GREE/) GREENE E A.  
Reeder RH, Moorefield B, Greene EA;  
WPI; 2002-681729/73.  
P-PSDB; ABB78308.  
Novel RRN3 polypeptide, an eukaryotic RNA polymerase I transcription factor and polynucleotide encoding it useful for diagnosing, treating hyper and hypoproliferative diseases in mammals and to identify modulators.



XX Claim 6; Page 25-26; 31pp; English.

PS The present sequence encodes a human RN3 polypeptide. RN3 is an

XX eukaryotic RNA polymerase I transcription factor. RN3 is useful for

CC screening for hypo or hyperproliferative diseases, including cancer,

CC malignancy, hyperplasia, metaplasia, dysplasia, benign tumour,

CC hyperproliferative disorder, benign dysplasia, benign tumour, autoimmune

CC disease or cardiac disease. It is also used to treat diseases involving

CC decreased cell proliferation, including degenerative disorders, growth

CC deficiencies, hypoproliferative disorders, physical trauma, lesions and

CC wounds. RN3 polypeptide, nucleic acid or antibody are also useful in

CC diagnostics, to detect, prognosis, diagnose or monitor various diseases.

CC RN3 nucleic acids are useful to identify other mammalian genes that

CC encode RN3-like molecules, to screen for mutations in a RN3 gene that

CC are associated with certain diseases. The polypeptide is useful as an

CC immunogen to generate antibodies which immunospecifically bind RN3

XX polypeptides

SQ Sequence 2068 BP; 573 A; 437 C; 485 G; 573 T; 0 U; 0 Other;

Query Match 100.0%; Score 2068; DB 6; Length 2068;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2068; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTGAGGCGCAGCGTGCCTTAGT 60

DB 1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTGAGGCGCAGCGTGCCTTAGT 60

QY 61 TCGGCCCAATGGCGGACCGCTGCTTCACACGGCTTTCGGGGAGATGCGGCGCTTCGT 120

DB 61 TCGGCCCAATGGCGGACCGCTGCTTCACACGGCTTTCGGGGAGATGCGGCGCTTCGT 120

QY 121 CCTCTGCAGTTAAGAGCTGGCGGCTCGAGGACTGGGATTTCAATATGCGTGCAATAG 180

DB 121 CCTCTGCAGTTAAGAGCTGGCGGCTCGAGGACTGGGATTTCAATATGCGTGCAATAG 180

QY 181 AGAATGACTTTTCAATTTCTCCCCAAGAAAACCTGTCGGTTTGGTGGAACTGTGACAG 240

DB 181 AGAATGACTTTTCAATTTCTCCCCAAGAAAACCTGTCGGTTTGGTGGAACTGTGACAG 240

QY 241 AAGTCTTGCTGAAGTACAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGC 300

DB 241 AAGTCTTGCTGAAGTACAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGC 300

QY 301 TGTAGATCCAGACATAAAGGATGACCCAGATCATCACTGGCTGCTAGATTTCCGTTCTT 360

DB 301 TGTAGATCCAGACATAAAGGATGACCCAGATCATCACTGGCTGCTAGATTTCCGTTCTT 360

QY 361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCCTT 420

DB 361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCCTT 420

QY 421 GGTGAATAGAAAGTCAACAGTAGTGGAGAGTATTTGGCTTTTCTTGGTAAATCTTGAT 480

DB 421 GGTGAATAGAAAGTCAACAGTAGTGGAGAGTATTTGGCTTTTCTTGGTAAATCTTGAT 480

QY 481 CAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATGCTTCCCATTTTGTGCCTC 540

DB 481 CAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATGCTTCCCATTTTGTGCCTC 540

QY 541 CCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTTCTGATGAAGATGATA 600

DB 541 CCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTTCTGATGAAGATGATA 600

QY 601 ATCTTCTCGAAATTTTGACACATGTACAGAGCCTTGCAATAATAGCAAGATATGTAC 660

DB 601 ATCTTCTCGAAATTTTGACACATGTACAGAGCCTTGCAATAATAGCAAGATATGTAC 660

QY 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAATTTCCATTTGTCGAAAAT 720

DB 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAATTTCCATTTGTCGAAAAT 720

QY 721 CAGAGAGAACACTGGAATGTTACGTTTACAACTTACTAAGGATTAGTGATATTTTCCAA 780

DB 721 CAGAGAGAACACTGGAATGTTACGTTTACAACTTACTAAGGATTAGTGATATTTTCCAA 780

QY 781 CCTTGAGGCATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTAATG 840

DB 781 CCTTGAGGCATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTAATG 840

QY 841 CATCCCGGAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGACAG 900

DB 841 CATCCCGGAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGACAG 900

QY 901 ATTCCACGGAAAGGATTTGTTAATATGGATGAAGATGAAGAACTGAACATGAACAAAGG 960

DB 901 ATTCCACGGAAAGGATTTGTTAATATGGATGAAGATGAAGAACTGAACATGAACAAAGG 960

QY 961 CTGCTCCTGAACGGCTCGACCAAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCCTGA 1020

DB 961 CTGCTCCTGAACGGCTCGACCAAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCCTGA 1020

QY 1021 TGTCTTTGGTTTTTGTCTTACATGAAGGATGTCTGTCTATGTAGATGGTAAAGTTGATAACG 1080

DB 1021 TGTCTTTGGTTTTTGTCTTACATGAAGGATGTCTGTCTATGTAGATGGTAAAGTTGATAACG 1080

QY 1081 GCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAAACCTCTGTTGCCCA 1140

DB 1081 GCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAAACCTCTGTTGCCCA 1140

QY 1141 CCCATGCTCTCCTGCCATGTACAGTTTTCATGTTTTTACCTCTGTAGTTTCAAAATGGGAT 1200

DB 1141 CCCATGCTCTCCTGCCATGTACAGTTTTCATGTTTTTACCTCTGTAGTTTCAAAATGGGAT 1200

QY 1201 TCGCAGAGGCATTTTGGAAACATCTCTGGAAAAAATTCAGAGGACCCCAAGTAATCCTGCCA 1260

DB 1201 TCGCAGAGGCATTTTGGAAACATCTCTGGAAAAAATTCAGAGGACCCCAAGTAATCCTGCCA 1260

QY 1261 TCATCAGGCAGGCTGCTGGAAAATTTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTC 1320

DB 1261 TCATCAGGCAGGCTGCTGGAAAATTTATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTC 1320

QY 1321 CTCTTATTACTGTAAATCATGCCATAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380

DB 1321 CTCTTATTACTGTAAATCATGCCATAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380

QY 1381 ATAAACCAGGATTCGGGAAACAAAGGCAATCTCGCATGTTGCTCTCCATGACCAATTTTACT 1440

DB 1381 ATAAACCAGGATTCGGGAAACAAAGGCAATCTCGCATGTTGCTCTCCATGACCAATTTTACT 1440

QY 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGGCGGAA 1500

DB 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGGCGGAA 1500

QY 1501 ACCTGAAAGAAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGGCGGATAGTATGAGCC 1560

DB 1501 ACCTGAAAGAAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGGCGGATAGTATGAGCC 1560

QY 1561 AGCTAAATCCCCCTGAAAGATTTGCCCTGCGCTCAGTGGTTAACTTTTGTGCAATCACAA 1620

DB 1561 AGCTAAATCCCCCTGAAAGATTTGCCCTGCGCTCAGTGGTTAACTTTTGTGCAATCACAA 1620

QY 1621 ATAAAGTACAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAAACAATCGCCAGATGCTGC 1680

DB 1621 ATAAAGTACAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAAACAATCGCCAGATGCTGC 1680

QY 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACA 1740

DB 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACA 1740

QY 1741 CCTTCTTCCCTTTTGTATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCCTATTT 1800

DB 1741 CCTTCTTCCCTTTTGTATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCCTATTT 1800

QY 1801 ATCAGGTATGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAA 1860



Db 1801 ATCAGGTATGGGAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAA 1860  
QY 1861 AGGACATAGTGGAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGATGATA 1920  
Db 1861 AGGACATAGTGGAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGATGATA 1920  
QY 1921 CCGTCATTTGGGATCACACCAAGCTCCTTTGACACGCTATTTCCGAAGTCTTCAAGTAGTG 1980  
Db 1921 CCGTCATTTGGGATCACACCAAGCTCCTTTGACACGCTATTTCCGAAGTCTTCAAGTAGTG 1980  
QY 1981 TGGGCTCCCCACCCGCTGTTGTACATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040  
Db 1981 TGGGCTCCCCACCCGCTGTTGTACATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040  
QY 2041 CTGAGATGTGACATTTGGGATTTCCCCAT 2068  
Db 2041 CTGAGATGTGACATTTGGGATTTCCCCAT 2068

RESULT 2  
AAA98384  
ID AAA98384 standard; cDNA; 2040 BP.  
XX  
AC AAA98384;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human RNA polymerase I transcription factor TIF-1A cDNA.  
XX  
KW RNA polymerase I transcription factor TIF-1A; antitumor; treatment;  
KW antiproliferative; cell proliferation; cancer; tissue regeneration; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200055316-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-DE000767.  
XX  
PR 17-MAR-1999; 99DE-01011992.  
XX  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
XX

Grummt I, Vingron M;  
WPI; 2000-587527/55.  
P-PSDB; AAB10936.  
New DNA encoding the transcription factor TIF-1A, useful for preventing or treating diseases associated with abnormal cell proliferation, particularly tumors.  
Claim 1; Fig 2; 38pp; German.

This invention describes a novel DNA sequence (I) that encodes the RNA polymerase I transcription factor TIF-1A which has antitumor, antiproliferative and proliferation-inducing activity. The invention also describes (1) DNA (Ia) encoding a protein (II) with the biological activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and able to bind specifically to, and cleave, its transcribed RNA so as to reduce or inhibit synthesis of the corresponding protein; (3) an antisense RNA (AS) with binding properties similar to R; (4) an expression vector that contains (I), (Ia) or sequences that encode R or AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II) encoded by (I) or (Ia); (7) preparation of TIF-1A or (II) by culturing cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists that weaken or block the activity of TIF-1A or (II); (10) a diagnostic method for detecting abnormal TIF-1A expression; and (11) kit for carrying out the method in (10). (I), and similar sequences that encode proteins with equivalent activity, expression vectors containing them, or their expression products are used to treat or prevent disorders

CC associated with reduced cellular proliferation, to stimulate cellular proliferation, and to promote tissue regeneration, e.g. after injury or radiation therapy. Ribozymes, antisense sequences directed against (I), also ligands and antagonists of TIF-1A, are used to treat or prevent disorders associated with excessive cellular proliferation and to inhibit proliferation, especially in treatment of cancers. (I) and specific ligands for TIF-1A (particularly antibodies (Ab)) are also useful for diagnosis of altered TIF-1A expression by (in)direct determination of the concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for detecting mutations. Ab can also be used for immunoprecipitation of TIF-1A and for isolation of related sequences from cDNA expression libraries. (I) allows recombinant production of TIF-1A in sufficient quantities for therapeutic use

SQ Sequence 2040 BP; 563 A; 433 C; 474 G; 568 T; 0 U; 2 Other;

Query Match 97.6%; Score 2018.4; DB 3; Length 2040;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2019; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 47 GCGGTGCGGTAGTTGCGGCCAATGGCGGACCGCTGCTTCACACGCGTTTGGCGGGAGA 106  
Db 18 GCGGTGCGGTAGTTGCGGCCAATGGCGGACCGCTGCTTCACACGCGTTTGGCGGGAGA 77  
QY 107 TCGGGCCGCTCGTCTCTGAGTTAAGAACTGGGCGCTGCGAGACTGGGATTTCAA 166  
Db 78 TCGGGCCGCTCGTCTCTGAGTTAAGAACTGGGCGCTGCGAGACTGGGATTTCAA 137  
QY 167 TATCGTGCATTAGAATGACTTTTCAATCTCCCCAAGAAAACTGTCGGTTGG 226  
Db 138 TATCGTGCATTAGAATGACTTTTCAATCTCCCCAAGAAAACTGTCGGTTGG 197  
QY 227 TGGAACTGTGACAGAAGTCTGCTGAAGTACAAAAAGGGTGAACAAATGACTTTGAGTT 286  
Db 198 TGGAACTGTGACAGAAGTCTGCTGAAGTACAAAAAGGGTGAACAAATGACTTTGAGTT 257  
QY 287 GTTGAAGAACCAAGCTGTTAGATCCAGACATAAAGGATGACCAAGATCATCAACTGGTGCT 346  
Db 258 GTTGAAGAACCAAGCTGTTAGATCCAGACATAAAGGATGACCAAGATCATCAACTGGTGCT 317  
QY 347 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAAGACTTTTGAGCAACTTATCAGTATTAT 406  
Db 318 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAAGACTTTTGAGCAACTTATCAGTATTAT 377  
QY 407 ATTAAGATTGCCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCT 466  
Db 378 ATTAAGATTGCCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCT 437  
QY 467 TGGTAATCTTGTATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATTGCTTC 526  
Db 438 TGGTAATCTTGTATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATTGCTTC 497  
QY 527 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTTCAGATTCTGA 586  
Db 498 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTTCAGATTCTGA 557  
QY 587 TGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTTCACAGAGCCTTGCATAAAT 646  
Db 558 TGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTTCACAGAGCCTTGCATAAAT 617  
QY 647 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACCTGGTGGAAAAATTTCC 706  
Db 618 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACCTGGTGGAAAAATTTCC 677  
QY 707 ATTTGTTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCATACTTACTAAGATTAG 766  
Db 678 ATTTGTTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCATACTTACTAAGATTAG 737  
QY 767 TGTATATTTTCCAACTTGGAGGATGAAATTTCTGGAGCTTATTATTGAAAAACTACTCAA 826  
Db 738 TGTATATTTTCCAACTTGGAGGATGAAATTTCTGGAGCTTATTATTGAAAAACTACTCAA 797  
QY 827 GTTGATGTGAATGCATCCCGGAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 886

Db	798	GTGGATGTGAATGCATCCCGCAGGGTATTGAGATGCTGAAGAAACAGCAACTCAAAC	857	
QY	887	TTGTGGTGGGACAGATCCACGGAAGGATTGTTTAATATGATGAAGATGAAGAACTGA	946	
Db	858	TTGTGGTGGGACAGATCCACGGAAGGATTGTTTAATATGATGAAGATGAAGAACTGA	917	
QY	947	ACATGAAACAAAGGCTGGTCCCTGAACGGCTCGACAGATGGTGATCCTGTAGCCGAGCG	1006	
Db	918	ACATGAAACAAAGGCTGGTCCCTGAACGGCTCGACAGATGGTGATCCTGTAGCCGAGCG	977	
QY	1007	CCTGGACATCCTGATGTCCTTGGTTTGTCTACATGAAGGATGTCGTATGTAGATGG	1066	
Db	978	CCTGGACATCCTGATGTCCTTGGTTTGTCTACATGAAGGATGTCGTATGTAGATGG	1037	
QY	1067	TAAGTGTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAA	1126	
Db	1038	TAAGTGTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAA	1097	
QY	1127	ACTCCTGTGCCACCCATGCCCTCCTGCCATGTACAGTCTTTTACGTCTCTGTAG	1186	
Db	1098	ACTCCTGTGCCACCCATGCCCTCCTGCCATGTACAGTCTTTTACGTCTCTGTAG	1157	
QY	1187	TTTCAAAATGGGATTCCGACAGGCAATTTTGGAAACATCTCTGGAAAAAATTCAGGACCC	1246	
Db	1158	TTTCAAAATGGGATTCCGACAGGCAATTTTGGAAACATCTCTGGAAAAAATTCAGGACCC	1217	
QY	1247	AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAG	1306	
Db	1218	AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAG	1277	
QY	1307	AGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCT	1366	
Db	1278	AGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCT	1337	
QY	1367	GCACATATACCTTAATAACCAAGGATTCGGGAAACAAAGGCATTCGCGATGTTGCTCTCCA	1426	
Db	1338	GCACATATACCTTAATAACCAAGGATTCGGGAAACAAAGGCATTCGCGATGTTGCTCTCCA	1397	
QY	1427	TGGACCAATTTTACTCAGCCTGCCAAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCA	1486	
Db	1398	TGGACCAATTTTACTCAGCCTGCCAAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCA	1457	
QY	1487	GCTTTTGGCGGAAACCTGAAAGAAAGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG	1546	
Db	1458	GCTTTTGGCGGAAACCTGAAAGAAAGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG	1517	
QY	1547	GATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCCTCAGTGGTTAACTTTTT	1606	
Db	1518	GATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCCTCAGTGGTTAACTTTTT	1577	
QY	1607	TGCTGCAATCACAATAAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA	1666	
Db	1578	TGCTGCAATCACAATAAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA	1637	
QY	1667	TCGCCAGATGCTGCCAGTCAATPAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC	1726	
Db	1638	TCGCCAGATGCTGCCAGTCAATPAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC	1697	
QY	1727	AAACCCGCTGGACACCTTCTTCCCTTTTGATCCCTGTGTGCTGAAGAGTCAAAGAAATT	1786	
Db	1698	AAACCCGCTGGACACCTTCTTCCCTTTTGATCCCTGTGTGCTGAAGAGTCAAAGAAATT	1757	
QY	1787	CATTGATCCTATTATCAGGTATGGGAAGACATGAGTGTCTGAAGCTACAGGAGTTCAA	1846	
Db	1758	CATTGATCCTATTATCAGGTGTGGGAAGACATGAGTGTCTGAAGCTACAGGAGTTCAA	1817	
QY	1847	GAAACCCATGAAAAAGGACATAGTGAAGATGAAGATGATGACTTTCTGAAGCGGAAGT	1906	
Db	1818	GAAACCCATGAAAAAGGACATAGTGAAGATGAAGATGATGACTTTCTGAAGCGGAAGT	1877	
QY	1907	GCCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCAATTTCCGAAG	1966	

Db

1878

GCCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCAATTTCCGAAG

1937

QY

1967

TCCTTCAAGTAGTGTGGCTCCCCACCCGTGTTGTATCATGCAACCCAGTCCCTCTTGACG

2026

Db

1938

TCCTTCAAGTAGTGTGGCTCCCCACCCGTGTTGTATCATGCAACCCAGTCCCTCTTGACG

1997

QY

2027

GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT

2068

Db

1998

GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT

2039

RESULT 3

ABS78724

ID

ABS78724

standard; DNA; 1770 BP.

XX

ABS78724;

DT

16-DEC-2002

(first entry)

XX

DNA encoding human NOVX3 protein.

XX

Human; NOVX; human disease; NOVX-associated disorder; cancer; addiction; Hodgkin disease; Von Hippel-Lindau syndrome; Parkinson's disease; stroke; tuberos sclerosis; hypercalcaemia; Parkinson's disease; depression; Huntington's disease; cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety; pain; obesity; Crohn's disease; osteoporosis; inflammatory bowel disease; infertility; inflammatory bowel disease; atherosclerosis; hypertension; scleroderma; haemophilia; diabetes; pancreatitis; autoimmune disease; asthma; arthritis; immunodeficiency; HIV; viral infection; neurogenesis; bacterial infection; parasitic infection; graft-versus-host disease; cell differentiation; cell proliferation; haematopoiesis; wound healing; angiogenesis; gene; ds.

XX

Homo sapiens.

OS

WO200272770-A2.

XX

19-SEP-2002.

XX

08-MAR-2002; 2002WO-US007283.

PF

08-MAR-2001; 2001US-0274281P.

XX

09-MAR-2001; 2001US-0274849P.

PR

12-MAR-2001; 2001US-0275235P.

PR

13-MAR-2001; 2001US-0275579P.

PR

13-MAR-2001; 2001US-0275601P.

PR

14-MAR-2001; 2001US-0276000P.

PR

20-MAR-2001; 2001US-0277239P.

PR

20-MAR-2001; 2001US-0277327P.

PR

20-MAR-2001; 2001US-0277338P.

PR

21-MAR-2001; 2001US-0277791P.

PR

22-MAR-2001; 2001US-0277833P.

PR

23-MAR-2001; 2001US-0278152P.

PR

26-MAR-2001; 2001US-0278894P.

PR

27-MAR-2001; 2001US-0279036P.

PR

28-MAR-2001; 2001US-0279344P.

PR

30-MAR-2001; 2001US-0280233P.

PR

02-APR-2001; 2001US-0280802P.

PR

02-MAY-2001; 2001US-0288148P.

PR

31-MAY-2001; 2001US-0294821P.

PR

31-OCT-2001; 2001US-0335302P.

PR

04-DEC-2001; 2001US-0338375P.

PR

07-MAR-2002; 2002US-00094466.

XX

(CURA-) CURAGEN CORP.

PA

Spytek KA, Vernet CA, Tchernev VT, Malyankar UM, Gerlach VL;

XX

Li L, Zethusen BD, Patturajan M, Gusev VY, Kekuda R, Pena CEA;

PI

Zhong M, Gangolli EA, Taupier RJ;

XX

WPI; 2002-713508/77.

DR

P-PSDB; ABG97480.

XX New NOVX polypeptides and polynucleotides, useful for preventing,  
PT diagnosing or treating NOVX-associated disorders, e.g. diabetes, multiple  
PT sclerosis, atherosclerosis, cancer, infections, osteoporosis or  
PT Parkinson's disease.  
XX

PS Claim 22; Page 98; 266pp; English.

XX The present invention relates to a new polypeptide (NOVX). The NOVX  
PS polypeptide, nucleic acid and antibody are useful in the manufacture of a  
CC medicament for treating a syndrome associated with a human disease,  
CC preferably a NOVX-associated disorder. The NOVX nucleic acids,  
CC polypeptides and antibodies are useful for treating, preventing or  
CC diagnosing diseases such as cancers, Hodgkin disease, Von Hippel-Lindau  
CC syndrome, Alzheimer's disease, stroke, tuberosus sclerosis,  
CC hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral  
CC palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-  
CC telangiectasia, leukodystrophies, addiction, anxiety, depression, pain,  
CC obesity, Crohn's disease, osteoporosis, inflammatory bowel disease,  
CC infertility, inflammatory bowel disease, atherosclerosis, hypertension,  
CC scleroderma, haemophilia, diabetes, pancreatitis, autoimmune disease,  
CC asthma, arthritis, immunodeficiencies, HIV, viral, bacterial or parasitic  
CC infections, or graft-versus-host disease. The nucleic acids and  
CC polypeptides may also be used as targets for the identification of small  
CC molecules that modulate or inhibit e.g. neurogenesis, cell  
CC differentiation, cell proliferation, haematopoiesis, wound healing and  
CC angiogenesis, in gene therapy, in generation of antibodies that bind  
CC immunospecifically to NOVX substances for use in therapeutic or  
CC diagnostic methods. The nucleic acids are further used as hybridisation  
CC probes, in chromosome mapping, tissue typing, preventive medicine, and  
CC pharmacogenomics. The present nucleic acid sequence encodes a human NOVX  
CC protein of the invention  
XX

SQ Sequence 1770 BP; 492 A; 373 C; 413 G; 492 T; 0 U; 0 Other;

Query Match 70.5%; Score 1457.6; DB 6; Length 1770;  
Best Local Similarity 86.2%; Pred. No. 0;  
Matches 1743; Conservative 0; Mismatches 9; Indels 270; Gaps 1;

QY 47 GCGGTCGGCTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 106  
DB 18 GCGGTCGGCTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 77  
QY 107 TCGCGCCGCTTCGTCCTCTGCAGTTAAGAGCTGGCGCTCGAGGACTGGGATTTCAA 166  
DB 78 TCGCGCCGCTTCGTCCTCTGCAGTTAAGAGCTGGCGCTCGAGGACTGGGATTTCAA 137  
QY 167 TATGCGTGCATTAGAGATGACTTTTCAATCTCCCCAAGAAACTGTTTCGGTTTG 226  
DB 138 TATGCGTGCATTAGAGATGACTTTTCAATCTCCCCAAGAAACTGTTTCGGTTTG 197  
QY 227 TGGAACTGTGACAGAGTCTTGCTGAAGTACAAAAGGCTGAACCAATGACTTTGAGTT 286  
DB 198 TGGAACTGTGACAGAGTCTTGCTGAAGTACAAAAGGCTGAACCAATGACTTTGAGTT 257  
QY 287 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAAGTCAATCAACTGGCTGCT 346  
DB 258 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAAGTCAATCAACTGGCTGCT 317  
QY 347 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 406  
DB 318 AGAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 377  
QY 407 ATTAAGATTGCCTTGGTTGAATAGAGTCAAAACAGTAGTGAAGAGTATTGGCTTTTCT 466  
DB 378 ATTAAGATTGCCTTGGTTGAATAGAGTCAAAACAGTAGTGAAGAGTATTGGCTTTTCT 437  
QY 467 TGGTAATCTTGATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATTGCTTTC 526  
DB 438 TGGTAATCTTGATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATTGCTTTC 497  
QY 527 CCATTTTGTGCCTCCCGAGTGATCAATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 586  
DB

DB 498 CCATTTTGTGCCTCCCGAGTGCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 557  
QY 587 TGATGAAGATGATAATCTTCCCTGCAAAATTTTGACACATGTACAGAGCCCTTGCAAAAT 646  
DB 558 TGATGAAGATGATAATCTTCCCTGCAAAATTTTGACACATGTACAGAGCCCTTGCAAAAT 617  
QY 647 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 706  
DB 618 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 677  
QY 707 ATTTGTTTCGAAAATCAGAGAGAACACTGGAAAGTTTACGTTTCAATACTTACTAAGGATTAG 766  
DB 678 ATTTGTTTCGAAAATCAGAGAGAACACTGGAAAGTTTACGTTTCAATACTTACTAAGGATTAG 737  
QY 767 TGTATATTTTCCAACTTGAGGCATGAAAATCTGGAGCTTATTTGAAAAACTACTCAA 826  
DB 738 TGTATATTTTCCAACTTGAGGCATGAAAATCTGGAGCTTATTTGAAAAACTACTCAA 797  
QY 827 GTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 886  
DB 798 GCTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 857  
QY 887 TTGTGGTGGGACAGATTCCACGGAAGGATTGTTAATATGGATGAAGATGAAGAAACTGA 946  
DB 858 TTGTGGTGGGACAGATTCCACGGAAGGATTGTTAATA----- 895  
QY 947 ACATGAAACAAAGGCTGGTCTCTGAACGCTCGACCAGATGCTGTCATCCTGTAGCCGAGCG 1006  
DB 896 ----- 895  
QY 1007 CCTGGACATCCTGATGTCTTTGGTTTGTCTCTACATGAAGGATGTCTGCTATGTAGTGG 1066  
DB 896 ----- 895  
QY 1067 TAAGGTTGATAAACGGCAAAACAAAGGATCTATATCGCGACTGTATAAACATCTTTGACAA 1126  
DB 896 ----- 895  
QY 1127 ACTCCTGTGCCACCCCATGCCCTCCTGCCATGTACAGTTTTCATGTTTTACCTCTGTAG 1186  
DB 896 ----- 895  
QY 1187 TTTCAAATTTGGGATTCGCAGAGGCATTTTGGAACATCTCTGGAAAAAATTCAGGACCC 1246  
DB 896 -----TGGGATTCGCAGAGGCATTTTGGAACATCTTTGGAAAAAATTCAGGATCC 947  
QY 1247 AAGTAATCTGCCATCATCAGGCGAGCTGCTGGAAATATATTGGAAGCTTTTGGCAAG 1306  
DB 948 AAGTAATCTGCCATCATCAGGCGAGCTGCTGGAAATATATTGGAAGCTTTTGGCAAG 1007  
QY 1307 AGCTAAATTTATTCCTCTTATTACTGTAAAAATCATGCTAGATCTTTTGGTTAACTGGCT 1366  
DB 1008 AGCTAAATTTATTTCTCTTATTACTGTAAAAACCATGCTAGATCTTTTGGTTAACTGGCT 1067  
QY 1367 GCACATATACCTTAATAAACCCAGGATTCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCA 1426  
DB 1068 GCACATATACCTTAATAAACCCAGGATTCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCA 1127  
QY 1427 TGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCA 1486  
DB 1128 TGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCA 1187  
QY 1487 GCTTTTGAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTTCAGAGTCTGAAATTTTGAGCG 1546  
DB 1188 GCTTTTGAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTTCAGAGTCTGAAATTTTGAGCG 1247  
QY 1547 GATAGTGATGAGCCAGCTAAATCCCTGAAGATTTCCTGAGTATCTTTCAGAGTCTGAAATTTTGAGCG 1606  
DB 1248 GATAGTGATGAGCCAGCTAAATCCCTGAAGATTTCCTGAGTATCTTTCAGAGTCTGAAATTTTGAGCG 1307  
QY 1607 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA 1666  
DB 1308 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAA 1367

QY 1667 TCGCCAGATGCTGCCAGTCAATTAGAGTACCGCTGGAGGAGACTCAGTGCAGATTCGAC 1726  
Db 1368 TCGCCAGATGCTGCCAGTCAATTAGAGTACCGCTGGAGGAGACTCAGTGCAGATTCGAC 1427  
QY 1727 AAACCCGCTGGACACCTTCTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATT 1786  
Db 1428 AAACCCGCTGGACACCTTCTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATT 1487  
QY 1787 CAITGATCCTATTTATCAGGTATGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAA 1846  
Db 1488 CAITGATCCTATTTATCAGGTGTGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAA 1547  
QY 1847 GAAACCCATGAAAAGGACATAGTGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGT 1906  
Db 1548 GAAACCCATGAAAAGGACATAGTGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGT 1607  
QY 1907 GCCCAGAAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCAATTTCCGAAG 1966  
Db 1608 GCCCAGAAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCAATTTCCGAAG 1667  
QY 1967 TCCTTCAAGTAGTGTGGCTCCCAACCGTGTGTGATGATGCAACCCAGTCCCTCTGACG 2026  
Db 1668 TCCTTCAAGTAGTGTGGCTCCCAACCGTGTGTGATGATGCAACCCAGTCCCTCTGACG 1727  
QY 2027 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCAT 2068  
Db 1728 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCAT 1769

RESULT 4  
ADD29697  
ID ADD29697 standard; mRNA; 1582 BP.  
XX  
AC ADD29697;  
XX  
XX 15-JAN-2004 (first entry)  
DT  
XX Human tumour suppressor mRNA SEQ ID NO:153.  
DE ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.  
XX  
XX Homo sapiens.  
XX WO2003058201-A2.  
XX  
XX 17-JUL-2003.  
XX  
XX 31-DEC-2002; 2002WO-US041825.  
PF  
XX 31-DEC-2001; 2001US-0345317P.  
XX  
XX (QUAR-) QUARK BIOTECH INC.  
PA (CLEV-) CLEVELAND CLINIC FOUND.  
XX  
PI Feinstein E, Gudkov AV;  
XX  
XX WPI; 2003-598393/56.  
DR  
XX  
XX Diagnosing cancer comprises determining the polypeptide or polynucleotide  
PT levels e.g., hepatic lipase, in a sample from a subject, where a higher  
PT level compared to that in a subject free of cancer is indicative of  
PT cancer.  
XX  
PS Disclosure; SEQ ID NO 153; 272pp; English.  
XX  
XX The invention relates to a novel method for diagnosing a cancer in a  
CC subject. the method comprises determining, in a sample from the subject,  
CC the level of at least one polypeptide, where a higher level of the  
CC polypeptide compared to the level of the polypeptide in a subject free of  
CC cancer is indicative of cancer. The polypeptide is selected from any of  
CC the polypeptides encoded by the polynucleotides listed in the  
CC specification and polypeptides which are at least 70% homologous to the

CC polypeptides. The method of the invention has cytostatic activity, and  
CC may have a use in gene therapy. The method is useful in identifying  
CC markers specific for one or several types of cancer, depending on the  
CC tissue origin, which may be used in numerous diagnostic and prognostic  
CC applications as well as cancer type-specific targets for therapeutic  
CC intervention. The compounds that modulate the activity of a tumour  
CC suppressor gene are useful in the treatment of cancer or as anti-cancer  
CC drugs. The present sequence represents a polynucleotide of the invention.  
XX  
SQ Sequence 1582 BP; 364 A; 253 C; 296 G; 389 T; 0 U; 280 Other;  
Query Match 60.6%; Score 1253.6; DB 9; Length 1582;  
Best Local Similarity 81.0%; Pred. No. 0;  
Matches 1276; Conservative 0; Mismatches 299; Indels 1; Gaps 1;  
QY 47 GCGGTGCGGTTAGTTCGGGCCCAATGGCGGCACCGCTGCTTCACACG-CGTTTCCCGGAG 105  
Db 1 GCGGTGCGGTTAGTTCGGGCCCAATGGCGGCACCGCTGCTTCACACGTTGTTGTCGGGAG 60  
QY 106 ATGCGGCGGCTTCGTCCTCTGCAAGTAAAGAGCTGGGCGCGTCGAGACTGGGATTTCAA 165  
Db 61 ATGCGGCGGCTTCGTCCTCTGCAAGTAAAGAGCGTGGGCGCGTCGAGACTGGGATTTCAA 120  
QY 166 ATATGCGTGCATTAGAGAATGACTTTTCAATTTCTCCCCCAAGAAACTGTTGCGGTTG 225  
Db 121 ATATGCGTGCATTAGAGAATGACTTTTCAATTTCTCCCCCAAGAAACTGTTGCGGTTG 180  
QY 226 GTGGAATGTTGACAGAAAGTCTTGTGAAAGTACAAAAGGGTGAACAAATGACTTTGAGT 285  
Db 181 GTGGAATGTTGACAGAAAGTCTTGTGAAAGTACAAAAGGGTGAACAAATGACTTTGAGT 240  
QY 286 TGTGGAAGAACCCAGCTGTTAGATCCAGACATAAAGAGTACCAGATCAACTGGCTGC 345  
Db 241 TGTGGAAGAACCCAGCTGTTAGATCCAGACATAAAGAGTACCAGATCAACTGGCTGC 300  
QY 346 TAGAATTCGGTTCCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTATCAGTATTA 405  
Db 301 TAGAATTCGGTTCCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTATCAGTATTA 360  
QY 406 TATTAGATTCGCTTGGTTGAATAGAAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTC 465  
Db 361 TATTAGATTCGCTTGGTTGAATAGAAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTC 420  
QY 466 TTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTT 525  
Db 421 TTGGTAATCTTGTATCAGCATAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTT 480  
QY 526 CCCATTTTGTGCTCCCGAGTGATCAATTAAGGAAGCGGATGTAGATGTTTCAGATTCG 585  
Db 481 CCCATTTTGTGCTCCCGAGTGATCAATTAAGGAAGCGGATGTAGATGTTTCAGATTCG 540  
QY 586 ATGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTACAGAGCCCTGCAAAATA 645  
Db 541 ATGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTACAGAGCCCTGCAAAATA 600  
QY 646 TAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAGAAATTTTC 705  
Db 601 TAGCAAGATATGTACCATCGACCGTGGTTTCTCATGCCAATACTGGTGAAGAAATTTTC 660  
QY 706 CATTTGTTGAAAATCAGAGAGAACACTGGAATGTTACGTTTATAAATTTACTAAGGATTA 765  
Db 661 CATTTGTTGAAAATCAGAGAGAACACTGGAATGTTACGTTTATAAATTTACTAAGGATTA 720  
QY 766 GTGTATATTTTCCAACTTGGAGGCATGAAATTTCTGGAGCTTATTTGAAAATACTACTCA 825  
Db 721 GTGTATATTTTCCAACTTGGAGGCATGAAATTTCTGGAGCTTATTTGAAAATACTACTCA 780  
QY 826 AGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAA 885  
Db 781 AGCTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAA 840  
QY 886 CTTGTGTTGGGACAGATTCCACGGGAAGGATTTTAAATATGATGAAGATGAAGAAACTG 945

Db 841 CTTGTGCTGGGACAGATTCCACGGAGGATGTTTAAATATGNNNNNNNNNNNNNNNNNN 900

QY 946 AACATGAAACAAAGGTGCTCTGAACGGCTCGACCAGATGGTGCACTGTAGCCGAGC 1005

Db 901 NNN 960

QY 1006 GCCTGGACATCCTGATGTCCTTTGGTTTGTCTCTACATGAAGGATGTCTGCTATGTAGATG 1065

Db 961 NNN 1020

QY 1066 GTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTTGACA 1125

Db 1021 NNN 1080

QY 1126 AACTCCTGTTGCCACCCCATGCCCTCCTGCCATGTACAGTTTTTTCATGTTTTTACCTCTGTA 1185

Db 1081 NNN 1140

QY 1186 GTTTCAAATTGGGATTCGCAGAGGCATTTTGGACATCTCTGGAACAAATTCGACGACC 1245

Db 1141 NNN 1200

QY 1246 CAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAA 1305

Db 1201 CAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAA 1260

QY 1306 GAGCTAAATTTAATCCTCTTTAATCTGTAATAATCATGCCTAGATCTTTTGGTTAACTGGC 1365

Db 1261 GAGCTAAATTTAATCCTCTTTAATCTGTAATAACCATGCCTAGATCTTTTGGTTAACTGGC 1320

QY 1366 TGCACATATACCTTAATAACCGAGATTCGGGAACAAAGGCATCTCGGATGTTGCTCTCC 1425

Db 1321 TGCACATATACCTTAATAACCGAGATTCGGGAACAAAGGCATCTCGGATGTTGCTCTCC 1380

QY 1426 ATGGACCATTTTACTCAGCCTGCGCAAGCTGTCTTCTACACCTTTGTTTTAGACACAAGC 1485

Db 1381 ATGGACCATTTTACTCAGCCTGCGCAAGCTGTCTTCTACACCTTTGTTTTAGACACAAGC 1440

QY 1486 AGCTTTTGGCGGAAACCTGAAAGAAGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGC 1545

Db 1441 AGCTTTTGGCGGAAACCTGAAAGAAGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGC 1500

QY 1546 GGATAGTGATGAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTT 1605

Db 1501 GGATAGTGATGAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTT 1560

QY 1606 TTGCTGCAATCACAAA 1621

Db 1561 TTGCTGCAATCACAAA 1576

RESULT 5

AAC76937

ID AAC76937 standard; cDNA; 1418 BP.

AC AAC76937;

XX

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4983.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiac;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive; ss.

XX Homo sapiens.

OS

XX WO200058473-A2.

PN

XX 05-OCT-2000.

PD

XX 31-MAR-2000; 2000WO-US008621.

PF

XX 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

PR 30-MAR-2000; 2000US-00540763.

XX

PA (CURA-) CURAGEN CORP.

XX

XX Shimkets RA, Leach M;

PI

XX WPI; 2000-602362/57.

DR P-PSDB; AAB42728.

DR

XX Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease.

PT

PT Claim 5; Page 4157-4158; 5507pp; English.

XX

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antidiabetic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 U; 0 Other;

Query Match 59.4%; Score 1228.4; DB 3; Length 1418;

Best Local Similarity 98.7%; Pred. No. 0;

Matches 1249; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 670 CGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTCCGAAATCAGAGAGAA 729

Db 1 CGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTCCGAAATCAGAGAGAA 60

QY 730 CACTGGAATGTTACGTTTCATACTACTAAGGATTAGTGTATATTTTCCAA-CCTTGAGG 788

Db 61 CACTGGAATGTTACGTTTCATACTACTAAGGATTAGTGTATATTTTCCAA-CCTTGAGG 120

QY 789 CATGAAATCTGGAGCTTATTATTGAAAAAATCTCAAGTTGGATGATGATCCCGG 848

Db 121 CATGAAATCTGGAGCTTATTATTGAAAAAATCTCAAGTTGGATGATGATCCCGG 180

QY 849 CAGGCTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTTGGTGGGACAGATTCCACG 908

Db 181 CAGGCTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTTGGTGGGACAGATTCCACG 240

QY 909 GAAGGATTGTTTAAATATGATGATGAAGAAACTGAACATGAACAAAGGCTGCTCT 968







		Query Match	48.9%;	Score 1010.4;	DB 4;	Length 1498;
		Best Local Similarity	98.8%;	Pred. No. 2.8e-289;		
		Matches 1028;	Conservative	0;	Mismatches 11;	Indels 1; Gaps 1;
QY	735	GAATGTTACGTTTCATAA	CTTACTAAAGGATTAGT	GTATATTTTCCAACCTTGAGGCATGAA	794	
Dd	4	GAAAGTTACGTTTCATAA	CTTACAGAGGATTAGT	GTATATTTTCCAACC-TGAGGCATGAA	62	
QY	795	ATTCTGGAGCTTA	TATTGAAAACTACTCAAGTTGGATGTGA	ATCGCATCCCGCAGGGT	854	
Dd	63	ATTCTGAGGCTTA	TATGAAAAAATACTCAAGTTGGATGTGA	ATCGCATCCCGCAGGGT	122	
QY	855	ATTGAAGATGCTGAAGAAA	CAGCAAACCTCAAACTTGTGGTGACAGATTCCACGGAAGGA	914		
Dd	123	ATTGAAGATGCTGAAGAAA	CAGCAAACCTCAAACTTGTGGTGACAGATTCCACGGAAGGA	182		
QY	915	TTGTTTTAATATGGA	TGAAGATGAAGAAAACCTGAACATGAAAACA	AAGGCTGGTCTCTGAACGG	974	
Dd	183	TTGTTTTAATATGGA	TGAAGATGAAGAAAACCTGAACATGAAAACA	AAGGCTGGTCTCTGAACGG	242	
QY	975	CTCGACCAGATGGTGC	ATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCTTTGGTTTTG	1034		
Dd	243	CTCGACCAGATGGTGC	ATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCTTTGGTTTTG	302		
QY	1035	TCCTACATGAAGGATGT	CTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACA	AAGGAT	1094	
Dd	303	TCCTACATGAAGGATGT	CTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACA	AAGGAT	362	
QY	1095	CTATATCGGACCTGATA	AAACATCTTTTGACAAACTCCTGTTGCCACCCATGCCTCCTGC	1154		
Dd	363	CTATATCGGACCTGATA	AAACATCTTTTGACAAACTCCTGTTGCCACCCATGCCTCCTGC	422		
QY	1155	CATGTACAGTTTTT	CATGTTTTACCTCTGTAGNTTCAAATTTGGANTTCGACAGGCA	1214		
Dd	423	CATGTACAGTTTTT	CATGTTTTACCTCTGTAGNTTCAAATTTGGANTTCGACAGGCA	482		
QY	1215	TTGGRAACATCTCTG	GAARAAAATTGCAGGACCCAAAGTAATCCTGCCATCATCAGGACGGCT	1274		
Dd	483	TTGGRAACATCTCTG	GAARAAAATTGCAGGACCCAAAGTAATCCTGCCATCATCAGGACGGCT	542		
QY	1275	GCTGGAAATTTAT	TTGGAAGCTTTTTTGGCAAGAGCTAAATTTATTCCTTTATTA	1334		
Dd	543	GCTGGAAATTTAT	TTGGAAGCTTTTTTGGCAAGAGCTAAATTTATTCCTTTATTA	602		
QY	1335	AAATCATGCTTAGATC	TTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCG	1394		
Dd	603	AAATCATGCTTAGATC	TTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCG	662		
QY	1395	GGAAACAAAGGCAT	TTCTGGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGCCAAGCT	1454		
Dd	663	GGAAACAAAGGCAT	TTCTGGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGCCAAGCT	722		
QY	1455	GTGTTCTACACCTT	TTGTTTATAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT	1514		
Dd	723	GTGTTCTACACCTT	TTGTTTATAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT	782		
QY	1515	TTGCAGTATCTTC	AGATCTGAAATTTTGAGCGGATAGTGTAGTGAGCCAGCTAAATCCCCTG	1574		
Dd	783	TTGCAGTATCTTC	AGATCTGAAATTTTGAGCGGATAGTGTAGTGAGCCAGCTAAATCCCCTG	842		
QY	1575	AAGATTTGCCTGCC	CTCAGTGGTTAACTTTTGTGCTCAATCACAAATAAGTACCAGCTC	1634		
Dd	843	AAGATTTGCCTGCC	CTCAGTGGTTAACTTTTGTGCTCAATCACAAATAAGTACCAGCTC	902		
QY	1635	GTCTTCTGCTAC	ACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTATTAGGAGT	1694		
Dd	903	GTCTTCTGCTAC	ACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTATTAGGAGT	962		
QY	1695	ACCGCTGGAGGAG	ACTCAGTGCAGATCTGCACAAAACCCGCTGGACACCTTCTTCCCCTTT	1754		
Dd	963	ACCGCTGGAGGAG	ACTCAGTGCAGATCTGCACAAAACCCCACTGGACACCTTCTTCCCCTTT	1022		

SQ	Sequence 1423 BP; 417 A; 247 C; 320 G; 439 T; 0 U; 0 Other;	
	Query Match 43.2%; Score 894; DB 4; Length 1423;	
	Best Local Similarity 80.9%; Pred. No. 1.2e-254;	
	Matches 1191; Conservative 0; Mismatches 10; Indels 271; Gaps 2;	
QY	157 GGATTCAAATAATGCGTGCATTAGAGAATGACTTTTCAATTCTCCCCAAGAAACTG 216	
Db		
QY	192 GGATTCAAATAATGCGTGCATTAGAGAATGATTTTTCAATTCTCCCCAAGAAACTG 251	
Db		
QY	217 TTCGGTTTGGTGGAACTGTGACAGAACTTGTCTGAAGTACAAAAAGGGTGAAACAATG 276	
Db		
QY	252 TTCGGTTTGGTGGAACTGTGACAGAACTTGTCTGAAGTACAAAAAGGGTGAAACAATG 311	
Db		
QY	277 ACTTTGAGTTGTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCA 336	
Db		
QY	312 ACTTTGAGTTGTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCA 371	
Db		
QY	337 ACTGGCTGTAGAAATTCGGTTCTTCTATCATGTACTTGAACAAAGACTTTGAGCAACTTA 396	
Db		
QY	372 ACTGGCTGTAGAAATTCGGTTCTTCTATCATGTACTTGAACAAAGACTTTGAGCAACTTA 431	
Db		
QY	397 TCAGTATTATATAAGATTGCCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATT 456	
Db		
QY	432 TCAGTATTATATAAGATTGCCCTTGGTTGAATAGAACTCAAAACAGTAGTGAAGAGTATT 491	
Db		
QY	457 TGGCTTTTCTTGGTAATCTTGATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCA 516	
Db		
QY	492 TGGCTTTTCTTGGTAATCTTGATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCA 551	
Db		
QY	517 TGATTGCTTCCCATTTTGTGCCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTT 576	
Db		
QY	552 TGATTGCTTCCCATTTTGTGCCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTT 611	
Db		
QY	577 CAGATTCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCT 636	
Db		
QY	612 CAGATTCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCT 671	
Db		
QY	637 TGCAAAATAAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGG 696	
Db		
QY	672 TGCAAAATAAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGG 731	
Db		
QY	697 AAAAAATTTCCATTTGTTCCGAAAAATCAGAGAGAACACTGGAAATGTTACGTTCAATACTTAC 756	
Db		
QY	732 AAAAAATTTCCATTTGTTCCGAAAAATCAGAGAGAACACTGGAAATGTTACGTTCAATACTTAC 791	
Db		
QY	757 TAAGGATTAGTGTATATTTTCCAACCTTGAGGCATGAAATCTGGAGCTTATTATTGAAA 816	
Db		
QY	792 TAAGGATTAGTGTATATTTTCCAACCTTGAGGCATGAAATCTGGAGCTTATTATTGAAA 851	
Db		
QY	817 AACTACTCAAGTTGGATGTAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAG 876	
Db		
QY	852 AACTACTCAAGTTGGATGTAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAG 911	
Db		
QY	877 CAACTCAAACTTGTGGTGGACAGATTCCACGGAAGGATTGTTTAATATGGATGAAGATG 936	
Db		
QY	912 CAAATCAAACTTGTGGTGGACAGATTCCACGGAAGGATTGTTTAATA----- 959	
QY	937 AAGAACTGAACATGAACAAAGGCTGGTCTGTAACGGCTCGACAGATGGTGCATCCTG 996	
Db	-----	
QY	960 TAGCCGAGCGCCTGGACATCCTGATGCTTTTGGTTTGTCTTACATGAAGGATGTCGCT 1056	
Db	-----	
QY	1057 ATGTAGATGTTAAGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACA 1116	
Db	-----	
QY	1117 TCCTTGACAAACTCCTGTTGCCCAACCCATGCCTCCTGCTCCATGTACAGTTTTTTCATGTTT 1176	
Db	-----	

QY	1177 ACCTCTGTAGTTTCAAATTTGGGATTCGCAGAGG-CATTTTTGGAAACATCTCTGGAAAAAA 1235	
Db		
QY	1236 TTGCAGGACCCAAAGTAATCCTGCCATCATCAGGAGGCTGCTGGAATATATTTGGAAGC 1295	
Db		
QY	1002 TTGCAGGATCCAAAGTAATCCTGCCATCATCAGGAGGCTGCTGGAATATATTTGGAAGC 1061	
Db		
QY	1296 TTTTGGCAAGAGCTAAAATTTATTCCTCTTATTACTGTAAATCATGCCTAGATCTTTTG 1355	
Db		
QY	1062 TTTTGGCAAGAGCTAAAATTTATTCCTCTTATTACTGTAAACCATGCCTAGATCTTTTG 1121	
Db		
QY	1356 GTTAACTGGCTGCACATATACCTTAATAACCCAGGATTCGGGAACAAAGGCATTTCTGCGAT 1415	
Db		
QY	1122 GTTAACTGGCTGCACATATACCTTAATAACCCAGGATTCGGGAACAAAGGCATTTCTGCGAT 1181	
Db		
QY	1416 GTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAGCTGTGTTCTACACCTTTGTTTTT 1475	
Db		
QY	1182 GTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTT 1241	
Db		
QY	1476 AGACACAAAGCAGCTTTTGGCGGAAACCTGAAAGAGTTTGCAGTATCTTTCAGAGTCTG 1535	
Db		
QY	1242 AGACACAAAGCAGCTTTTGGCGGAAACCTGAAAGAGTTTGCAGTATCTTTCAGAGTCTG 1301	
Db		
QY	1536 AATTTTGGCGGATAGTGATGAGCCAGCTAAATCCCTGAAAGATTTGCTGCCCTCAGTG 1595	
Db		
QY	1302 AATTTTGGCGGATAGTGATGAGCCAGCTAAATCCCTGAAAGATTTGCTGCCCTCAGTG 1361	
Db		
QY	1596 GTTAACTTTTGTGCAATCACAAATAAGTA 1627	
Db		
QY	1362 GTTAACTTTTGTGCAATCACAAAGATGAA 1393	
Db		
RESULT 8		
AAS92255		
ID	AAS92255 standard; cDNA; 2410 BP.	
XX		
AC	AAS92255;	
XX		
DT	13-FEB-2002 (first entry)	
XX		
DE	DNA encoding novel human diagnostic protein #28059.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US008631.	
XX		
PR	31-MAR-2000; 2000US-00540217.	
PR	23-AUG-2000; 2000US-00649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Drmanac RT, Liu C, Tang YT;	
XX		
DR	WPI; 2001-639362/73.	
DR	P-PSDB; ABG28068.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity.	
XX		
PS	Claim 1; SEQ ID NO 28059; 103pp; English.	
XX		
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)	

sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 U; 0 Other;

Query Match 30.9%; Score 639.4; DB 5; Length 2410;  
Best Local Similarity 96.8%; Pred. No. 8.9e-179;  
Matches 706; Conservative 0; Mismatches 16; Indels 7; Gaps 5;

QY	1347	GATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCGGGAACAAAGGCA	1406
Db	1	GATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCGGGAACAAAGGCA	60
QY	1407	TTCTGCGATGTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACC	1466
Db	61	TTCTGCGATGTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACC	120
QY	1467	TTTGTTTTACACACAAGCAGCTTTTGGCGGAAACCTGAAGAAGGTTTGCGAGTATCTT	1526
Db	121	TTTGTTTTACACACAAGCAGCTTTTGGCGGAAACCTGAAGAAGGTTTGCGAGTATCTT	180
QY	1527	CAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGAT-TTGCCCT	1585
Db	181	CAGAGTCTGAATTTTGAGTGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATATGGCTT	240
QY	1586	GCCCTCAGTGGTTAACTTTTGTGCTGCAATCACAATAAAGTACCAGCTCGTCTTCTGCTA	1645
Db	241	GCCCTCAGTGGTTAACTTTTGTGCTGCAATCACAATAAAGTACCAGCTCGTCTTCTGCTA	300
QY	1646	CACCATCATGAGAGGAACAATCGCCAGATGCTGCCAGTGCTATTAGGAGTACCGCTGGAGG	1705
Db	301	CACCATCATGAGAGGAACAATCGCCAGATGCTGCCAGTGCTATTAGGAGTACCACTGGAGG	360
QY	1706	AGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTCCCTTTT-GATCCCTGTG	1764
Db	361	AGACTCAGTGCAGACCTGCACAAACCCACTGGACACCTTCTCCCTTTTGGATCCCTGTG	420
QY	1765	TGCTGAAGAGGTCAAAGAAATT-CATTGATCCTATT---ATCAGGTATGGGAAGACATG	1820
Db	421	TGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTTATCCAGGTATGGGAAGACATG	480
QY	1821	AGTGTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAGGACATAGTGAAGATGAA	1880
Db	481	AGTGTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAGGACATAGTGAAGATGAA	540
QY	1881	GATGATGATTTCTGAAAGCGAAGTGCCCGAGAATGATACCGTGATTGGGATCACACCA	1940
Db	541	GATGATGATTTCTGAAAGCGAAGTGCCCGAGAATGATACCGTGATTGGGATCACACCA	600
QY	1941	AGCTCCTTTGACACGCATTTCCGAAGTCTTCAAGTAGTGTGGGCTCCCCACCCGTGTG	2000
Db	601	AGCTCCTTTGACACGCATTTCCGAAGTCTTCAAGTAGTGTGGGCTCCCCACCCGTGTG	660
QY	2001	TACA-TGCAACCCAGTCCCTCTGTACGCGCAGAAATTTGTGACTGAGATGTGACATTGGG	2059

Db	661	TACACTGCAACCCAGTCCCCTCTGACGGCAGAAATTTGCGACTGAGATGTGACATTGGG	720
QY	2060	ATTCCCCCAT 2068	
Db	721	ATTCCCCCAT 729	
RESULT 9			
AAS92253			
ID	AAS92253	standard; cDNA; 776 BP.	
XX	AC	AAS92253;	
XX	DT	13-FEB-2002 (first entry)	
XX	DE	DNA encoding novel human diagnostic protein #28057.	
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX	OS	Homo sapiens.	
XX	PN	WO200175067-A2.	
XX	PD	11-OCT-2001.	
XX	PF	30-MAR-2001; 2001WO-US008631.	
XX	PR	31-MAR-2000; 2000US-00540217.	
XX	PR	23-AUG-2000; 2000US-00649167.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Drmanac RT, Liu C, Tang YT;	
XX	DR	WPI; 2001-639362/73.	
XX	DR	P-PSDB; ABG28066.	
XX	PT	New isolated polynucleotide and encoded polypeptides, useful in	
XX	PT	diagnostics, forensics, gene mapping, identification of mutations	
XX	PT	responsible for genetic disorders or other traits and to assess	
XX	PT	biodiversity.	
XX	PS	Claim 1; SEQ ID NO 28057; 103pp; English.	
XX	CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)	
XX	CC	sequences. (I) is useful as hybridisation probes, polymerase chain	
XX	CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,	
XX	CC	and in recombinant production of (II). The polynucleotides are also used	
XX	CC	in diagnostics as expressed sequence tags for identifying expressed	
XX	CC	genes. (I) is useful in gene therapy techniques to restore normal	
XX	CC	activity of (II) or to treat disease states involving (II). (II) is	
XX	CC	useful for generating antibodies against it, detecting or quantitating a	
XX	CC	polypeptide in tissue, as molecular weight markers and as a food	
XX	CC	supplement. (II) and its binding partners are useful in medical imaging	
XX	CC	of sites expressing (II). (I) and (II) are useful for treating disorders	
XX	CC	involving aberrant protein expression or biological activity. The	
XX	CC	polypeptide and polynucleotide sequences have applications in	
XX	CC	diagnostics, forensics, gene mapping, identification of mutations	
XX	CC	responsible for genetic disorders or other traits to assess biodiversity	
XX	CC	and to produce other types of data and products dependent on DNA and	
XX	CC	amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic	
XX	CC	coding sequences of the invention. Note: The sequence data for this	
XX	CC	patent did not appear in the printed specification, but was obtained in	
XX	CC	electronic format directly from WIPO at	
XX	CC	ftp.wipo.int/pub/published_pct_sequences	
SQ	Sequence 776 BP; 206 A; 157 C; 198 G; 215 T; 0 U; 0 Other;		

Query Match 27.5%; Score 568.6; DB 5; Length 776;  
Best Local Similarity 97.8%; Pred. No. 5.2e-158;  
Matches 629; Conservative 0; Mismatches 9; Indels 5; Gaps 5;

QY 4 GAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTCAGCGCGCAGCGGTCCGCTTAGTTCG 63  
Db |||||||  
QY 64 GCCAATGGCGGCACCGCTGCTTTCACAGC-CGTTTGCCGGGAGATGCGGCGCTTCCTCC 122  
Db |||||||  
QY 193 GCCCAATGGCGGCACCGCTGCTTTCACAGCTGTTTGTGCGGAGATGCGGCGCTTCCTCC 252  
Db |||||||  
QY 123 TCTGCAGTTAAGAAGCTGGCGCGCTCG-AGGACTGGGANTTCAAATATGCGTGCAATAGA 181  
Db |||||||  
QY 253 TCTGCAGTCAAGACGCTGTTTCGCGCTGTAGGACTGGGANTTCAAATATGCGTGCAATAGA 312  
QY 182 GAATGACTTTTCAATTTCTCCCCCAAGAAAAAAGTTCGGTTCGGTGAAGTGTGACAGA 241  
Db |||||||  
QY 313 GAATGATTTTTCATTTCTCCCCCAAGAAAAAAGTTCGGTTCGGTGAAGTGTGACAGA 372  
QY 242 AGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGCT 301  
Db |||||||  
QY 373 AGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGCT 432  
QY 302 GTTAGATCCAGACATAAAGGATGA-CCAGATCATCAACT-GGCTGTGAGAAATTCGTTCT 359  
Db |||||||  
QY 433 GTTAGATCCAGACATAAAGGATGACCCAGATCATCAACTGGGCTGCTAGAAATTCGTTCT 492  
QY 360 TCTATCATGTACTTGCACAAAAGACTTTGAGCAACTTATCAGTATTATTAAGATTGCC- 418  
Db |||||||  
QY 493 TCTGTCTGTACTTTGACAAAAAGACTTTGAGCAACTTATCAGTATTATTAAGATTGCC 552  
QY 419 TTGGTTGAATAGAGTCAAAACAGTAGTGGAAAGAGTATTGGCTTTTCTGGTAATCTTGT 478  
Db |||||||  
QY 553 TTGGTTGAATAGAGTCAAAACAGTAGTGGAAAGAGTATTGGCTTTTCTGGTAATCTTGT 612  
QY 479 ATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTCCTCCCATTTTGTGCC 538  
Db |||||||  
QY 613 ATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTCCTCCCATTTTGTGCC 672  
QY 539 TCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCGTGATGAAGATGA 598  
Db |||||||  
QY 673 TCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCGTGATGAAGATGA 732  
QY 599 TAATCTTCCCTGCAAAATTTTGACACATGTCAAGAGCCCTTGCAA 641  
Db |||||||  
QY 733 TAATCTTCCCTGCAAAATTTTGACACATGTCAAGAGCCCTTGCAA 775

RESULT 10  
ABS64829  
ID ABS64829 standard; DNA; 2493 BP.

XX AC ABS64829;

XX DT 15-NOV-2002 (first entry)

XX DE Human tumour suppressor gene #3.

XX KW Human; cancer; gene; ds; tumour suppressor gene; breast cancer; lymphoma; adenocarcinoma; leukaemia; melanoma; myeloma; sarcoma; teratocarcinoma; prostate; cervix; liver; ovary; adrenal gland; heart; brain; lung; colon; placenta; skeletal muscle; synovial membrane; tonsil; kidney; uterus; skin; cytostatic.

XX OS Homo sapiens.

XX PN WO200264775-A1.

XX PD 22-AUG-2002.

XX PF 12-FEB-2002; 2002WO-AU000137.

XX PR 12-FEB-2001; 2001AU-00003052.

XX PR 12-FEB-2001; 2001AU-00003053.

XX PR 12-FEB-2001; 2001AU-00003054.

PA (BION-) BIONOMICS LTD.  
XX Callen DF, Powell JA, Kremmidiotis G, Gardner AE, Whitmore SA;  
XX WPI; 2002-657597/70.  
DR P-PSDB; ABG79517.  
XX New tumor suppressor genes identified at 16q24.3, useful for identifying or obtaining full-length human genes involved in the tumorigenic process, or in diagnosing or treating cancer (e.g. breast or cervix cancer) through gene therapy.  
PS Claim 10; Page 76-79; 218pp; English.  
XX The invention relates to a tumour suppressor gene and the protein it encodes. The genes are useful for identifying and/or obtaining full-length human genes involved in the tumorigenic process. The genes the polypeptides and antibodies to the polypeptides are useful in diagnosing cancer, in establishing the prognosis of a patient diagnosed with cancer, or in treating cancer through gene therapy. In particular, the cancer is breast cancer. The drugs that restore tumour suppressor activity are useful for treating breast cancer or for manufacturing a medicament for the treatment of breast cancer. The cancer may also be adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, or cancer of the prostate, cervix, liver, ovary, adrenal gland, heart, brain, lung, placenta, skeletal muscle, synovial membrane, tonsil, kidney, colon, uterus or skin. This sequence represents a human tumour suppressor gene of the invention  
SQ Sequence 2493 BP; 554 A; 700 C; 761 G; 478 T; 0 U; 0 Other;  
Query Match 27.4%; Score 567; DB 6; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 3.1e-157;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1196 GGGATTGCGCAGAGGCAATTTTGGAAACATCTCTGGAAAAAATTCAGAGCAACCAAGTAATCC 1255  
Db |||||||  
QY 711 GGGATTGCGCAGAGGCAATTTTGGAAACATCTCTGGAAAAAATTCAGAGCAACCAAGTAATCC 770  
QY 1256 TGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGAGCTAAATT 1315  
Db |||||||  
QY 771 TGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGAGCTAAATT 830  
QY 1316 TATTCTCTTATTACTGTAAATCATGCTAGATCTTTGGTTAACTGGCTGCACATATA 1375  
Db |||||||  
QY 831 TATTCTCTTATTACTGTAAATCATGCTAGATCTTTGGTTAACTGGCTGCACATATA 890  
QY 1376 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCATGGACCAAT 1435  
Db |||||||  
QY 891 CCTTAATAACCAAGGATTCGGGAACAAAGGCATTTCTGCAATGTTGCTCTCCATGGACCAAT 950  
QY 1436 TTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAAGCAGCTTTTGAG 1495  
Db |||||||  
QY 951 TTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAAGCAGCTTTTGAG 1010  
QY 1496 CGGAAACCTGAAAGAGGTTTGCAATCTTCAGAGTCTGAAATTTTGAGCGGATAGTGAT 1555  
Db |||||||  
QY 1011 CGGAAACCTGAAAGAGGTTTGCAATCTTCAGAGTCTGAAATTTTGAGCGGATAGTGAT 1070  
QY 1556 GAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTAACTTTTGTGCAAT 1615  
Db |||||||  
QY 1071 GAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTAACTTTTGTGCAAT 1130  
QY 1616 CACAAATAGTACCAGCTCGTCTTCTGTACACCATCATATTGAGAGGAACAATCGCCAGAT 1675  
Db |||||||  
QY 1131 CACAAATAGTACCAGCTCGTCTTCTGTACACCATCATCGAGAGGAACAATCGCCAGAT 1190  
QY 1676 GCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCT 1735  
Db |||||||  
QY 1191 GCTGCCAGTCATTAGGAGTACCACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCCACT 1250  
QY 1736 GGACACCTTCTTCCCTTTTGATCCCTGTGTGCTGAAGAGGTCA 1778  
|||

Db	1251	GGACACCTTCGTCCCTTTGATCCCTGTGTGCTGAAGAGGTGA	1293
RESULT 11			
ADA52976			
ID	ADA52976	standard; cDNA; 2493 BP.	
XX			
AC	ADA52976;		
XX			
DT	20-NOV-2003	(first entry)	
XX			
DE	Human coding sequence, SEQ ID 544.		
XX			
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;		
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;		
KW	inflammatory disease; osteoporosis; neurological disease; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	EP1293569-A2.		
XX			
PD	19-MAR-2003.		
XX			
PF	21-MAR-2002; 2002EP-00006586.		
XX			
PR	14-SEP-2001; 2001JP-00328381.		
PR	24-JAN-2002; 2002US-0350435P.		
XX			
PA	(HELI-) HELIX RES INST.		
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.		
XX			
PI	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;		
PI	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;		
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;		
XX			
DR	WPI; 2003-395539/38.		
DR	P-PSDB; ADA54615.		
XX			
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory		
PT	and/or membrane proteins, useful for developing medicines for diseases in		
PT	which the gene is involved, or as target molecules for gene therapy.		
XX			
PS	Claim 1; SEQ ID NO 544; 205pp; English.		
XX			
CC	The present invention relates to novel human secretory or membrane		
CC	proteins (ADA54072-ADA5710) and their coding sequences (ADA52433-		
CC	ADA54071). The coding sequences are useful in the gene therapy of		
CC	diseases caused by abnormalities of the proteins, e.g. cancer,		
CC	inflammatory diseases, osteoporosis or neurological disease.		
XX			
SQ	Sequence 2493 BP; 554 A; 700 C; 761 G; 478 T; 0 U; 0 Other;		
	Query Match 27.4%; Score 567; DB 7; Length 2493;		
	Best Local Similarity 98.3%; Pred. No. 3.1e-157;		
	Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;		
QY	1196	GGGATTGCGAGAGGCATTTTGGAAACATCTCTGGAAAAAATTGCAGGACCAAGTAATCC	1255
Db	711	GGGATTGCGAGAGGCATTTTGGAAACATCTCTGGAAAAAATTGCAGGATCCAAGTAATCC	770
QY	1256	TGCCATCATCAGGCAGGCTGCTGGAAATTATATGGAGCTTTTGGCAAGAGCTAAATT	1315
Db	771	TGCCATCATCAGGCAGGCTGCTGGAAATTATATGGAGCTTTTGGCAAGAGCTAAATT	830
QY	1316	TATTCTCTTATTACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGTGCACATATA	1375
Db	831	TATTCTCTTATTACTGTAAACCATGCCTAGATCTTTTGGTTAACTGGTGCACATATA	890
QY	1376	CCTTAATACAGGATTTCGGAAACAAAGGCATTCTGGATTTGCTCTCCATGGACCAATT	1435
Db	891	CCTTAATACAGGATTTCGGAAACAAAGGCATTCTGCAATGTGCTCTCCATGGACCAATT	950
QY	1436	TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGAGCTTTTGAG	1495
Db	951	TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAG	1010
QY	1496	CGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAAATTTTGAGCGGATAGTGAT	1555
Db	1011	CGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAAATTTTGAGCGGATAGTGAT	1070
QY	1556	GAGCCAGCTAAATCCCTGAAGATTTCCTGCCCTCAGTGGTTAACTTTTGTGCTGCAAT	1615
Db	1071	GAGCCAGCTAAATCCCTGAAGATTTCCTGCCCTCAGTGGTTAACTTTTGTGCTGCAAT	1130
QY	1616	CACAAATAAGTACCAGCTCGTCTTCTGTCTACACCATCATTTGAGAGGAACAATCGCCAGAT	1675
Db	1131	CACAAATAAGTACCAGCTCGTCTTCTGTCTACACCATCATCGAGAGGAACAATCGCCAGAT	1190
QY	1676	GCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCT	1735
Db	1191	GCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGACTGCACAAACCCACT	1250
QY	1736	GGACACCTTCTTCCCTTTTGATCCCTGTGTGCTGAAGAGGTCA	1778
Db	1251	GGACACCTTCGTCCCTTTGATCCCTGTGTGCTGAAGAGGTGA	1293
RESULT 12			
AAS87643			
ID	AAS87643	standard; cDNA; 1260 BP.	
XX			
AC	AAS87643;		
XX			
DT	13-FEB-2002	(first entry)	
XX			
DE	DNA encoding novel human diagnostic protein #23447.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US0008631.		
XX			
PR	31-MAR-2000; 2000US-00540217.		
PR	23-AUG-2000; 2000US-00649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
XX	P-PSDB; ABG23456.		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity.		
XX			
PS	Claim 1; SEQ ID NO 23447; 103pp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)		
CC	sequences. (I) is useful as hybridisation probes, polymerase chain		
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,		
CC	and in recombinant production of (II). The polynucleotides are also used		
CC	in diagnostics as expressed sequence tags for identifying expressed		
CC	genes. (I) is useful in gene therapy techniques to restore normal		
CC	activity of (II) or to treat disease states involving (II). (II) is		
CC	useful for generating antibodies against it, detecting or quantitating a		
CC	polypeptide in tissue, as molecular weight markers and as a food		
CC	supplement. (II) and its binding partners are useful in medical imaging		
CC	of sites expressing (II). (I) and (II) are useful for treating disorders		



CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1260 BP; 344 A; 227 C; 269 G; 420 T; 0 U; 0 Other;  
  
Query Match 26.6%; Score 550.8; DB 5; Length 1260;  
Best Local Similarity 89.0%; Pred. No. 1.4e-152;  
Matches 696; Conservative 0; Mismatches 77; Indels 9; Gaps 9;  
  
QY 4 GAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTGAGCGCAGCGGTCCGTTAGTTTCG 63  
Db |||||  
133 GAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTGAGCGCAGCGGTCCGTTAGTTTCG 192  
QY 64 GCCCAATGGCGGCACCGCTGCTTCACACG-CGTTTCCGGGAGATGCGGCGCTTCGTCC 122  
Db |||||  
193 GCCCAATGGCGGCACCGCTGCTTCACACGTTGTTGTGCGGAGATGCGGCGCTTCGTCC 252  
QY 123 TCTGCAGTTAAGAAAGCTGGCGCGCTG-AGGACTGGGATTCAAATATGGTGCATTAGA 181  
Db |||||  
253 TCTGCAGTCAAGACGCTGTTGCGCTGCTAGGACTGGGATTCAAATATGGTGCATTAGA 312  
QY 182 GAATGACTTTTCAATTCTCCCCCAAGAAAAACTGTTCCGTTTGGTGGAACTGTGACAGA 241  
Db |||||  
313 GAATGATTTTCAATTCTCCCCCAAGAAAAACTGTTCCGTTTGGTGGAACTGTGACAGA 372  
QY 242 AGTCTTGCTGAAGTACAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCGACT 301  
Db |||||  
373 AGTCTTGCTGAAGTACAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCGACT 432  
QY 302 GTTAGATCCAGACATAAAGGATGA-CCAGATCATCAACT-GGCTGCTAGAAATCCGTTCT 359  
Db |||||  
433 GTTAGATCCAGACATAAAGGATGACCCAGATCATCAACTGGGCTGCTAGAAATCCGTTCT 492  
QY 360 TCTATCATGTACTTGA-CAAAAGACTTTGAG-CAACTATCAGTATTATATTAAGATTGC 417  
Db |||||  
493 TCTGTCACTGACTTGACCAAAAGACTTTGAGCCAACTTATCAGTATTATATGAGATTGC 552  
QY 418 C-TTGGTTGAATAGAACTC-AAACAGTAGTGGAAAGATTTGGCTTTTCTTGGTAATCT 475  
Db |||||  
553 CTTTGGTTGAATAGAACTC-AAACAGTAGTGGAAAGATTTGGCTTTTCTTGGTAATCT 612  
QY 476 TGTATCAGCAGACACTGTTTCTCAGACCGTGTCTCAGCATGATTGCTT-CCCATTTTG 534  
Db |||||  
613 TGTATCAGCAGACACTGTTTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTG 672  
QY 535 TGCCTCCCGGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGATGATGAAG 594  
Db |||||  
673 TGCCTCCCGGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGATGATGAAG 732  
QY 595 ATGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCCTGCAAAATATAGCAAGAT 654  
Db |||||  
733 ATGATAATCTTCTGCAAAATTTTGACACATATACAGAGCCCTGCAAAATATAGCAAGAT 792  
QY 655 ATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTC 714  
Db |||||  
793 ATGTACCATCATGAGTATACCTTTCTTATTTGAAATGTTTAATTTCTCAAGAAATTTGA 852  
QY 715 GAAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAACTTACTAAGGATTAAGTATATT 774  
Db |||||  
853 ATCAATTAGTAAAAATTATAAAATGTTAATAGTATTAAAGCTTGAGTCTTACATTGCATT 912  
QY 775 TT 776  
Db ||  
913 TT 914

RESULT 13  
AAS80955/c  
ID AAS80955 standard; cDNA; 655 BP.  
XX  
AC AAS80955;  
XX 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #16759.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US0008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
(HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG16768.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 16759; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 655 BP; 207 A; 134 C; 115 G; 199 T; 0 U; 0 Other;

Query Match 23.1%; Score 477.8; DB 5; Length 655;  
Best Local Similarity 90.2%; Pred. No. 4.8e-131;  
Matches 578; Conservative 0; Mismatches 57; Indels 6; Gaps 6;  
  
QY 108 GCGGCCGCTTCGTCTCTGCAGTTAAGAAGCTGGGCGCGTCG-AGGACTGGGATTTCAA 166  
Db |||||  
655 GCGGCCGCTTCGTCTCTGCAGTCAAGACGCTGTTTCGCGTCTAGGACTGGGATTTCAA 596  
QY 167 TATGCGTGCATTAGAGAATGACTTTTTCAAATTTCTCCCCAAGAAAACTGTTCCGTTTGG 226  
Db |||||



Db 595 TATGCGTGCAATTAGAGAAATGATTTTTCATTCTCCCCCAAGAAAACTGTTTCGGTTTG 536

QY 227 TGGAACGTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAAAACAATGACTTTGAGTT 286

Db 535 TGGAACGTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAAAACAATGACTTTGAGTT 476

QY 287 GTTGAAGAACCAAGCTGTTAGATCCAGACATATAAAGGATGA-CCAGATCATCAACTGGCTGC 345

Db 475 GTTGAAGAACCAAGCTGTTAGATCCAGACATATAAAGGATGACCCAGATCATCAACTGGCTGC 416

QY 346 TAGAATCCGTTCTTCTATCATGTACTTGA-CAAAAGACTTTTGAG-CAACTTATCAGTAT 403

Db 415 TAGAATCCGTTCTTCTATCATGTACTTGA-CAAAAGACTTTTGAG-CAACTTATCAGTAT 356

QY 404 TATATTAAGATTGCTTGGTTGTAATPAGAAAGTC-AAACAGTAGTGGAAGAGTATTTGGCTT 462

Db 355 TATATTGAGATTGCTTGGTTGTAATPAGAAAGTC-AAACAGTAGTGGAAGAGTATTTGGCTT 296

QY 463 TTCTTTGGTAATCTTGTTATCAGCAGACTGTTTTCCTCAGACCCGTGCTCAGCATGATTG 522

Db 295 TTCTTTGGTAATCTTGTTATCAGCAGACTGTTTTCCTCAGACCCGTGCTCAGCATGATTG 236

QY 523 CTT-CCCATTTTGTGCCTCCCGAGTGATCAATTAAGGAAGGCGATGTAGATGTTTCAGAT 581

Db 235 CTTCCCATTTTGTGCCTCCCGAGTGATCAATTAAGGAAGGCGATGTAGATGTTTCAGAT 176

QY 582 TCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAA 641

Db 175 TCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAA 116

QY 642 ATAATAGCAAGATATGTACCATCGACACCCGTGTTTCTCATGCCAATACTGGTGGAATA 701

Db 115 ATAATAGCAAGATATGTACCATCATGAGTATATCTTTTCCTTATTTTGAATGTTAATTCT 56

QY 702 TTTCCATTGTTTCGAAAAATCAGAGAGAACACTGGAATGTTA 742

Db 55 CAAGAAAAATTGTAATCAATTAGTAAAAAATTATAAAATGTTA 15

RESULT 14

AAS87641

ID AAS87641 standard; cDNA; 3169 BP.

XX

AC AAS87641;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #23445.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR P-PSDB; ABG23454.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

Claim 1; SEQ ID NO 23445; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 U; 0 Other;

Query Match 22.5%; Score 466.2; DB 5; Length 3169;

Best Local Similarity 64.6%; Pred. No. 3.5e-127;

Matches 933; Conservative 0; Mismatches 213; Indels 299; Gaps 4;

QY 69 ATGGCGGCACCGCTGCTTCACACGCGTTTGCGGGGAGATGCGCGCTTCGCTCTGCA 128

Db 1 ATGGCGGCACCGCTGCTTCACACGCGTTTGCGGGGAGATGCGCGCTTCGCTCTGCA 60

QY 129 GTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCAATAGAGAATGAC 188

Db 61 GTCAAGATGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCAATAGAGAATGAT 120

QY 189 TTTTTCAAATTTCCCCCAAGAAAAAATGTTTCGGTTTGTTGGAACCTGTGACAGAAGTCTTG 248

Db 121 TTTTTCAAATTTCCCCCAAGAAAAAATGTTTCAGTTTGGTGAACCTGTGACAGAAGTCTTG 180

QY 249 CTGAAGTACAAAAAGGGTGAAAAACAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT 308

Db 181 CTGAAGTACAAAAAGGGTGAAAAACAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT 240

QY 309 CCAGACATAAAGGATGACCCAGATCATCAACTGGCTGCTAGAAATTCGGTTCTTCTATCATG 368

Db 241 CCAGACATAAAGGATGCTGTGCCCGCATGCCATGATCTCTGAGGAGGGGAGGGCCATT 300

QY 369 TACTTGACAAAAAGACTTTGAGCAACTTATCAGTATTAATAAGATTGCCTTGGTTGAAT 428

Db 301 GTGGGCGAGTGTGCCCTTTCATCCCTTCGTGGCTTTCAGTTGTCTCAAACAGAGTGGC 360

QY 429 AGAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAAATCTTGTATCAGCACAG 488

Db 361 CTCAGCCGGAAGTTGCTAAGCTG-----CCTGCGGTGGTTCGG 399

QY 489 ACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCGAGTG 548

Db 400 CCGGTGGCTGCCGCGCCACATACAGCCCCAGCCTCTGCCACAACCCCAACGCCACCG 459

QY 549 ATCATTAAGGAAGGCGATGTAGATGTTTCAGATTTCTGATGATGAAGATGATAATCTTCTT 608

Db 460 CAGCTTCTGGCTCACTCATCTGCTTGCA----- 488

QY 609 GCAAAATTTTGACACATGTACAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGACA 668

Db 489 -----GGACA 493

QY 669 CCGTGGTTTCTCATGCCAATACTGGTGAAAAAATTTCCATTTGTTGAAAAATCAGAGAGA 728

Db 494 CCGTGGTTCTCATGCAATACTGGTGAAAAATTTCCATTGTTCCGAAAATCAGAGAGA 553  
QY 729 ACACCTGGAATGTTACGTTTACAACTTAAAGGATTAGTGATATTTTCCAAACCTTGAGG 788  
Db 554 ACACCTGGAATGTTACGTTTACAACTTAAAGGATTAGTGATATTTTCCAAACCTTGAGG 613  
QY 789 CATGAATTTCTGGAGCTTATTATTGAAAACTACTCAAGTTGGATGTAATGCATCCCGG 848  
Db 614 CATGAATTTCTGGAGCTTATTATTGAAAACTACTCAAGTTGGATGTAATGCATCCCGG 673  
QY 849 CAGGCTATTGAAGATGCTGAAGAAACAGCAACTCAAACCTTGTGGTGGACAGATTCCACG 908  
Db 674 CAGGCTATTGAAGATGCTGAAGAAACAGCAAAATCAAACCTTGTGGTGGACAGATTCCACG 733  
QY 909 GAAGGATTGTTTAAATATGGATGAAGATGAAGAAACTGAACATGAACAAAGGCTGTCTT 968  
Db 734 GAAGGATTGTGTAATATG----- 752  
QY 969 GAACGGCTGACCATGATGGTGATCCTCTAGCCGAGCGCCTGGACATCCTGTATCTTTG 1028  
Db 753 ----- 752  
QY 1029 GTTTTGTCTACATGAAGGATGCTGCTATGTAGTGAAGTTGTATAACGGCAAAACA 1088  
Db 753 --TTAGCAGTTTATTAAATGAAGTGGAGATGAAGTTTATCATAATCAAAAGGTGGAACA 810  
QY 1089 AAGGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTTGGCCACCCATGCC 1148  
Db 811 G----- 811  
QY 1149 TCCTGCCATGTACAGTTTTTTCATGTTTAACTCTGTAGTTTCAAATTTGGGATTCGCAGAG 1208  
Db 812 -----CTAGTCTGTCTCATCTTTGGATTCGCAGAG 841  
QY 1209 GCATTTTGGAAACATCTCTGGAAAAAATTCAGGACCCCAAGTAATCCTGCCATCATCAGG 1268  
Db 842 GCATTTTGGAAACCTCTTTGGAAAAAATTCAGGATCCCAAGTAATCCTGCCATCATCAGG 901  
QY 1269 CAGGCTGCTGGAAATTATATTTGGAAAGCTTTTGGCAAGAGCTAAATTTTATCTCTTATT 1328  
Db 902 CAGGCTGCTGGAAATTATATTTGGAAAGCTTTTGGCAAGAGCTAAATTTATCTCTTATT 961  
QY 1329 ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAG 1388  
Db 962 ACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAG 1021  
QY 1389 GATTCCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC 1448  
Db 1022 GATTCCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGC 1081  
QY 1449 CAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAGCAGCTTTTGGCGGAAACCTGAAA 1508  
Db 1082 CAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAGCAGCTTTTGGCGGAAACCTGAAA 1141  
QY 1509 GAAGG 1513  
Db 1142 GAAGG 1146

RESULT 15  
ABA83038  
ID ABA83038 standard; DNA; 1461 BP.  
XX AC ABA83038;  
XX DT 05-FEB-2002 (first entry)  
XX DE Human transcription factor TRFX-65 coding sequence.  
KW Human; transcription factor; TRFX; cell proliferative disease;  
KW autoimmune disease; inflammation; neurological disease;  
KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;

KW neuroprotective; antiinflammatory; gene therapy; ds.  
OS Homo sapiens.  
PN WO200172777-A2.  
PD 04-OCT-2001.  
PF 13-MAR-2001; 2001WO-US008117.  
XX 13-MAR-2000; 2000US-0188986P.  
PR (INCY-) INCYTE GENOMICS INC.  
PA Hillman JL, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;  
PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;  
PI Reddy R;  
XX WPI; 2001-570896/64.  
DR P-PSDB; ABB50214.  
XX Novel transcription factor polypeptides, used to treat diseases  
PT associated with altered activity and expression of TRFX, and to screen  
PT for agents capable of modulating its activity.  
XX Claim 11; Page 299; 327pp; English.  
XX The present sequence is the coding sequence for a human transcription  
CC factor. The transcription factor and its coding sequence are useful in  
CC the diagnosis, treatment and prevention of diseases associated with  
CC altered expression of the transcription factor e.g. cell proliferative,  
CC autoimmune/inflammatory, neurological and developmental disorders. A  
CC number of specific disorders/diseases are given in the specification,  
CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
CC disease, stroke, and viral, bacterial, fungal and protozoal infections  
XX  
SQ Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 U; 0 Other;  
Query Match 20.3%; Score 420.8; DB 5; Length 1461;  
Best Local Similarity 97.3%; Pred. No. 7.1e-114;  
Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 1188 TTCAAATTTGGGATTTCGACAGGCGCATTTTGGAAACATCTCTGGAAAAAATTCGAGGACCCA 1247  
Db 530 TTAAATATGGGATTTCGACAGGCGCATTTTGGAAACATCTCTGGAAAAAATTCGAGGATCCA 589  
QY 1248 AGTAATCCTGCCATCATCAGGCGAGGCTGCTGGAAATTAATTTGGAGCTTTTGGCAAGA 1307  
Db 590 AGTAATCCTGCCATCATCAGGCGAGGCTGCTGGAAATTAATTTGGAGCTTTTGGCAAGA 649  
QY 1308 GCTAAATTTATTCTCTTATTACTGTAAATCATGCCTAGATCTTTTGGTTAACTGGCTG 1367  
Db 650 GCTAAATTTATTCTCTTATTACTGTAAACCATGCCTAGATCTTTTGGTTAACTGGCTG 709  
QY 1368 CACATATACCTTAATAACCGAGGATTCGGGAACAAAGGCATTTCTCGATGTTGCTCTCCAT 1427  
Db 710 CACATATACCTTAATAACCGAGGATTCGGGAACAAAGGCATTTCTCGATGTTGCTCTCCAT 769  
QY 1428 GGACCATTTTACTCAGCCTGCCAAGCTGTGTCTACACCTTTGTTTTAGACACAAGCAG 1487  
Db 770 GGACCATTTTACTCAGCCTGCCAAGCTGTGTCTACACCTTTGTTTTAGACACAAGCAG 829  
QY 1488 CTTTTGAGCGGAAACCTGAAAGAGGTTTTCAGATATCTTCAGAGTCTGAATTTTGAGCGG 1547  
Db 830 CTTTTGAGCGGAAACCTGAAAGAGGTTTTCAGATATCTTCAGAGTCTGAATTTTGAGCGG 889  
QY 1548 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTT 1607

---

Db	890	ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTGGCCTGCCCTCAGTGGTTAACTTTTT	949
Qy	1608	GCTGCAATCACAAATAAGTA	1627
Db	950	GCTGCAATCACAAAGATGAA	969

Search completed: July 18, 2004, 02:37:31  
Job time : 827 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 02:18:21 ; Search time 153 Seconds  
(without alignments)  
7500.905 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagaggctgtggctgaag.....tgacatttgggattcccat 2068

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	208.2	10.1	513	4	US-09-621-976-3086
2	48.2	2.3	8797	2	US-08-723-306-6
3	48.2	2.3	8797	5	PCT-US96-10041-6
4	48.2	2.3	11093	2	US-08-723-306-5
5	48.2	2.3	11093	5	PCT-US96-10041-5
C 6	43.4	2.1	399	4	US-09-621-976-8976
C 7	40.4	2.0	7218	1	US-08-232-463-14
8	39.2	1.9	3108	4	US-09-206-942-72
9	39.2	1.9	4937	1	US-08-038-682-3
10	39.2	1.9	4937	1	US-08-302-832-3
11	39.2	1.9	4937	2	US-08-530-198-3
12	39.2	1.9	4937	2	US-08-469-880-3
13	39.2	1.9	4937	2	US-08-728-470-3
14	39.2	1.9	4937	2	US-08-617-697-3
15	39.2	1.9	4937	3	US-08-719-641-3
16	39.2	1.9	4937	4	US-09-206-942-70
C 17	38	1.8	832	4	US-09-621-976-2813
C 18	37.8	1.8	2418	4	US-09-601-198-61
19	37.8	1.8	2910	4	US-09-206-942-31
20	37.8	1.8	2928	4	US-09-206-942-29
21	37.2	1.8	7518	4	US-09-620-312D-1051
22	36.6	1.8	1497	4	US-09-220-132-94
23	36.2	1.8	399	4	US-09-621-976-8976
24	36.2	1.8	586	4	US-09-669-751-55
25	35.8	1.7	832	4	US-09-621-976-2813
26	35.6	1.7	1398	4	US-09-134-001C-1019
C 27	35.6	1.7	2148	4	US-09-219-983A-19

C 28	35.6	1.7	2492	4	US-09-219-983A-18	Sequence 18, Appl
29	35.4	1.7	1461	4	US-09-369-247-49	Sequence 49, Appl
30	35.2	1.7	5538	2	US-08-231-193A-55	Sequence 55, Appl
31	35.2	1.7	5538	2	US-08-486-273A-55	Sequence 55, Appl
32	35.2	1.7	5538	3	US-08-940-086A-55	Sequence 55, Appl
33	35.2	1.7	5538	4	US-08-940-035A-55	Sequence 55, Appl
34	35.2	1.7	5538	4	US-08-935-105A-55	Sequence 55, Appl
35	35.2	1.7	5538	4	US-09-648-797-55	Sequence 55, Appl
36	35.2	1.7	5538	4	US-09-386-123-55	Sequence 55, Appl
37	35.2	1.7	5983	4	US-08-264-578-1	Sequence 1, Appli
38	35.2	1.7	10878	4	US-09-911-842A-1	Sequence 1, Appli
39	35	1.7	3757	4	US-09-620-312D-106	Sequence 106, App
C 40	35	1.7	1230025	4	US-09-198-452A-1	Sequence 1, Appli
41	34.8	1.7	304	4	US-09-313-294A-5811	Sequence 5811, Ap
42	34.8	1.7	771	4	US-09-134-001C-2374	Sequence 2374, Ap
43	34.8	1.7	978	4	US-09-134-001C-714	Sequence 714, App
44	34.6	1.7	2518	3	US-09-433-699-3	Sequence 3, Appli
45	34.6	1.7	3222	4	US-09-206-942-48	Sequence 48, Appl

ALIGNMENTS

RESULT 1  
US-09-621-976-3086  
; Sequence 3086, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3086  
; LENGTH: 513  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 25..231  
US-09-621-976-3086

Query Match	10.1%;	Score 208.2;	DB 4;	Length 513;
Best Local Similarity	98.6%;	Pred. No. 5.9e-54;		
Matches 210;	Conservative	0; Mismatches	3; Indels	0; Gaps 0;
QY 1650	ATCATTGAGGAGGAACAATCGCCAGATGCTGCCAGTATAGGAGTACCGCTGGAGGAGAC	1709		
Db 1	ATCATTGAGGAGGAACAATCGCCAGATGCTGCCAGTATAGGAGTACCGCTGGAGGAGAC	60		
QY 1710	TCAGTGCAGATCTGCACAAACCCGGTGGACACCTTCTTCCCTTTGATCCCTGTGTGCTG	1769		
Db 61	TCAGTGCAGACCTGCACAAACCCACTGGACACCTTCTTCCCTTTGATCCCTGTGTGCTG	120		
QY 1770	AAGAGTCAAAGAAATTCATTGATCCTATTATCAGGTATGGGAAGACATGAGTGTGCTGAA	1829		
Db 121	AAGAGTCAAAGAAATTCATTGATCCTATTATCAGGTATGGGAAGACATGAGTGTGCTGAA	180		
QY 1830	GAGTACAGGAGTTCAAGAAACCCATGAAAAAG	1862		
Db 181	GAGTACAGGAGTTCAAGAAACCCATGAAAAAG	213		

RESULT 2  
US-08-723-306-6  
; Sequence 6, Application US/08723306  
; Patent No. 5856178  
; GENERAL INFORMATION:  
; APPLICANT: White PhD, Kenneth



APPLICANT: Morrey PhD, John  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic  
Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City  
STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/723,306  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweigert PhD, Susan E  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Construct comprising Bos  
DESCRIPTION: taurus beta casein 5' regulatory region plus genes encoding  
DESCRIPTION: amphipathic peptide and green fluorescent protein"  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
US-08-723-306-6

Query Match 2.3%; Score 48.2; DB 2; Length 8797;  
Best Local Similarity 52.2%; Pred. No. 0.00083;  
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 697 AAAAATTTCCATTTGTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAACTTAC 756  
Db 2278 AAAGAAGAGTATATTATTAAATTTGCTCAGAACATCCAAATTTCAAGTTTATCATTAT 2337  
QY 757 TAAGGATTAGTGTATATTTCACCTTGAGGCGATGAAATTTCTGGAGCTTATTATTGAAA 816  
Db 2338 CTTACAATATTTCAAAAATATTAAATATAGATACATGAAATACAGAAAGTAAATTAAGAGA 2397  
QY 817 AACTACTCAAGTTGGATGTGAATGCATCCCGGCGAGGATGTTGAAGATGCTGAAGAAACAG 876  
Db 2398 AAGTATTTTATTTTGTAATAAATTTCTAGGTTGGACAGGGAGTACCAGGAAACAAAA 2457  
QY 877 CAACTCAAACTTGTGGTGGGACAGA 901  
Db 2458 ACAATGAAAAATGTGATCTGACAGA 2482

RESULT 3  
PCT-US96-10041-6  
Sequence 6, Application PC/TUS9610041  
GENERAL INFORMATION:  
APPLICANT: White PhD, Kenneth  
APPLICANT: Morrey PhD, John  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic  
Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City  
STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10041  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweigert PhD, Susan E  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Construct comprising Bos  
DESCRIPTION: taurus beta casein 5' regulatory region plus genes encoding  
DESCRIPTION: amphipathic peptide and green fluorescent protein"  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
PCT-US96-10041-6

Query Match 2.3%; Score 48.2; DB 5; Length 8797;  
Best Local Similarity 52.2%; Pred. No. 0.00083;  
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 697 AAAAATTTCCATTTGTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAACTTAC 756  
Db 2278 AAAGAAGAGTATATTATTAAATTTGCTCAGAACATCCAAATTTCAAGTTTATCATTAT 2337  
QY 757 TAAGGATTAGTGTATATTTCACCTTGAGGCGATGAAATTTCTGGAGCTTATTATTGAAA 816  
Db 2338 CTTACAATATTTCAAAAATATTAAATATAGATACATGAAATACAGAAAGTAAATTAAGAGA 2397  
QY 817 AACTACTCAAGTTGGATGTGAATGCATCCCGGCGAGGATGTTGAAGATGCTGAAGAAACAG 876  
Db 2398 AAGTATTTTATTTTGTAATAAATTTCTAGGTTGGACAGGGAGTACCAGGAAACAAAA 2457  
QY 877 CAACTCAAACTTGTGGTGGGACAGA 901  
Db 2458 ACAATGAAAAATGTGATCTGACAGA 2482

RESULT 4  
US-08-723-306-5  
Sequence 5, Application US/08723306  
Patent No. 5856178  
GENERAL INFORMATION:  
APPLICANT: White PhD, Kenneth  
APPLICANT: Morrey PhD, John  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic  
Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City

STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/723,306  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweigert PhD, Susan E  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11093 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Construct comprising portions of Bos taurus beta casein gene and genes encoding amphipathic peptide and green fluorescent protein"  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1801..1834  
OTHER INFORMATION: /product= "beta casein exon 1"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3780..3832  
OTHER INFORMATION: /product= "beta casein exon 2"  
FEATURE:  
NAME/KEY: TATA signal  
LOCATION: 1766..1773  
FEATURE:  
NAME/KEY: exon  
LOCATION: 4567..4590  
OTHER INFORMATION: /product= "portion of beta casein exon 3"  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 127..1800  
OTHER INFORMATION: /function= "5' flanking regulatory region of bovine beta casein gene"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 4587..5310  
OTHER INFORMATION: /product= "Green fluorescent protein"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 5320..5449  
OTHER INFORMATION: /product= "Shiva-1 coding sequence"  
FEATURE:  
NAME/KEY: polyA signal  
LOCATION: 7630..7635  
OTHER INFORMATION: /standard\_name= "Bovine beta casein exon 9"  
OTHER INFORMATION: 3' region, in exon 9"

US-08-723-306-5

Query Match 2.3%; Score 48.2; DB 2; Length 11093;  
Best local Similarity 52.2%; Pred. No. 0.00096;  
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 697 AAAAATTCATTTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAACTTAC 756

Db 2278 AAAGAAGAGTATATTTTAAATTTGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337  
QY 757 TAAGGATTAGTGATATTTTCCAACTTGGAGCATGAAATTTCTGGAGCTTATTATTGAAA 816  
Db 2338 CTTACAATATTTCAAAAATATTAATATAGATACATGAAATACAGAAAGTAAATTAAGAGA 2397  
QY 817 AACTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAGATGCTGAAGAAACAG 876  
Db 2398 AAGTATTTTATTTGTAATAAAAAAATTTCTAGTTGGACAGGGAGTACCAGGAACAAAAA 2457  
QY 877 CAACTCAAACCTTGTGGTGGGACAGA 901  
Db 2458 ACAATGAAAAATGTGATCTGACAGA 2482

RESULT 5  
PCT-US96-10041-5  
Sequence 5, Application PC/TUS9610041  
GENERAL INFORMATION:  
APPLICANT: White PhD, Kenneth  
APPLICANT: Morrey PhD, John  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City  
STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10041  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweigert PhD, Susan E  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11093 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Construct comprising portions of Bos taurus beta casein gene and genes encoding amphipathic peptide and green fluorescent protein"  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1801..1834  
OTHER INFORMATION: /product= "beta casein exon 1"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3780..3832  
OTHER INFORMATION: /product= "beta casein exon 2"  
FEATURE:  
NAME/KEY: TATA signal  
LOCATION: 1766..1773  
FEATURE:  
NAME/KEY: exon  
LOCATION: 4567..4590  
OTHER INFORMATION: /product= "portion of beta casein exon 3"  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 127..1800  
OTHER INFORMATION: /function= "5' flanking regulatory region of bovine beta casein gene"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 4587..5310  
OTHER INFORMATION: /product= "Green fluorescent protein"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 5320..5449  
OTHER INFORMATION: /product= "Shiva-1 coding sequence"  
FEATURE:  
NAME/KEY: polyA signal  
LOCATION: 7630..7635  
OTHER INFORMATION: /standard\_name= "Bovine beta casein exon 9"  
OTHER INFORMATION: 3' region, in exon 9"

NAME/KEY: exon  
LOCATION: 4567..4590  
OTHER INFORMATION: /product= "portion of beta casein  
OTHER INFORMATION: exon 3"  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 127..1800  
OTHER INFORMATION: /function= "5' flanking regulatory  
OTHER INFORMATION: region of bovine beta casein gene"  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 4587..5310  
OTHER INFORMATION: /product= "Green fluorescent  
OTHER INFORMATION: protein"  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 5320..5449  
OTHER INFORMATION: /product= "Shiva-1 coding sequence"  
FEATURE:  
NAME/KEY: polyA signal  
LOCATION: 7630..7635  
OTHER INFORMATION: /standard\_name= "Bovine beta casein  
OTHER INFORMATION: 3' region, in exon 9"  
PCT-US96-10041-5

Query Match 2.3%; Score 48.2; DB 5; Length 11093;  
Best Local Similarity 52.2%; Pred. No. 0.00096;  
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 697 AAAAAATTCATTTGTCGAAAATCAGAGAGAACTGGAATGTTACGTTTCATAACTTAC 756  
Db 2278 AAAGAGAGATATATTTTAAATGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337  
QY 757 TAAGGATTAGTGATATATTTCCAACTTGAGGCATGAAATTTCTGGAGCTTATTTGAAA 816  
Db 2338 CTTACAATATTTTCAAAAATATTAATAATAGATACATGAAATACAGAAATAAAGAGA 2397  
QY 817 AACTACTCAAGTTGGATGTAATGCATCCCGGAGGGTATTGAAGATGCTGAAGAAACAG 876  
Db 2398 AAGTATTTTATTTGTAAAAAATAATTTCTAGTTGGACAGGAGTACCAGGAACAAAAA 2457  
QY 877 CAACTCAAACCTTGCTGGGGACAGA 901  
Db 2458 ACAATGAAAAATGTGATCTGACAGA 2482

RESULT 6  
US-09-621-976-8976/c  
; Sequence 8976, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 8976  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-8976

Query Match 2.1%; Score 43.4; DB 4; Length 399;  
Best Local Similarity 11.8%; Pred. No. 0.0035;  
Matches 35; Conservative 138; Mismatches 124; Indels 0; Gaps 0;  
QY 895 GGACAGATTCCACGGAAGGATGTTTAATATGGATGAAGATGAAGAACTGAACATGAAA 954  
Db 330 SRGSYRRYKTSAMMGRANKMKRKKTKMYWMKGGGKGGSTYMAMRSRRGSTGRWSYRRA 271

QY 955 CAAAGGCTGCTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGCGCTGGACA 1014  
Db 270 MWRGSKSWGGSYYRMAGYRSSRWRSWYSAMWRKKKMTCKGRSSWGRSTGYAWMYKK 211  
QY 1015 TCCTGATGCTTTGGTTTGTCTACATGAAGGATGCTGCTATGTAGATGGTAAGGTTG 1074  
Db 210 SWCTSRKWMYKKRRKKWRKCTSTKRTCYRGSTYKCKAYYTKRRKRWTRWYTYKSY 151  
QY 1075 ATAAAGGCAAAACAAAGGATCTATATCGGACCTGATAAACATCTTTGACAAACTCCTGT 1134  
Db 150 MSMKKTWRMKTAYYWTXKWKMTRTKWTCTMCKCTTYMMAGTMMYRVRYYVYAKRAKW 91  
QY 1135 TGCCCAACCCATGCTCCTGCCATGTACAGTCTTTTACCTCTGTAGTTTCA 1191  
Db 90 SKRCTWSTTCYCMKYMAKKCWSYWWSMSMMKWGKSMWKWTYYYYYMMKWSKMTYW 34

RESULT 7  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 2.0%; Score 40.4; DB 1; Length 7218;  
Best Local Similarity 6.8%; Pred. No. 0.19;  
Matches 29; Conservative 207; Mismatches 188; Indels 0; Gaps 0;  
QY 584 TGATGATGAAGATGATAATCTTCTCGCAAATTTTGACACATGTACAGAGCCTTGCAAAT 643

Db 1458 TAAAGAGATAGAAGAAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1399

QY 644 AATAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATATCGTGGAAAAATT 703

Db 1398 RRR 1339

QY 704 TCCATTGTTGAAATCAGAGAGAACACTGGAATGTTACGTTTATTAACATTACTAAGGAT 763

Db 1338 RRR 1279

QY 764 TAGTGTATATTTTCCAACTTGAGGCAATGAAATCTCGAGCTTATTATTGAAAAACTACT 823

Db 1278 RRR 1219

QY 824 CAAGTTGGATGTGAATGCATCCCGGCAGGTATTGAAGATGCTGAAGAAACAGCAACTCA 883

Db 1218 RRR 1159

QY 884 AACTTGTGGTGGACAGATCCACGGAAGGATTGTTTAAATATGGATGAAGTGAAGAAAC 943

Db 1158 RRR 1099

QY 944 TGAACATGAAACAAAGCGTGGTCTCTGAACGGCTCGACAGATGGTGCATCCTGTAGCCGA 1003

Db 1098 RRR 1039

QY 1004 GCGC 1007

Db 1038 GCTC 1035

RESULT 8

US-09-206-942-72

; Sequence 72, Application US/09206942

; Patent No. 6432669

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High

; FILE REFERENCE: 1038-861 MIS:jb

; CURRENT APPLICATION NUMBER: US/09/206,942

; EARLIER FILING DATE: 1998-12-08

; EARLIER FILING DATE: 1998-10-07

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 72

; LENGTH: 3108

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

US-09-206-942-72

Query Match 1.9%; Score 39.2; DB 4; Length 3108;

Best Local Similarity 53.2%; Pred. No. 0.26;

Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTGAAAAAACTACTCAAGTTGGATGTGAATGCATCCCGCAGGCTATTGAAGA 862

Db 2613 GGTATTAAACGCAAAAGATGCTAAGCTAAATGGTGTATGCATCAGGTATACAGAAGT 2672

QY 863 TGCTGAAGAAACAGCAACTAAACTTGTGTGGGACAGATCCACGGAAGGATTGTTAA 922

Db 2673 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGA 2732

QY 923 TATGGATGAAGATGAAGAAACTGAACATGAACAAAA 958

Db 2733 TATCACTGGGGATTAAACACAGTAAATGGGTAAA 2768

RESULT 9

US-08-038-682-3

; Sequence 3, Application US/08038682

; Patent No. 5549897

; GENERAL INFORMATION:

; APPLICANT: BARENKAMP, STEPHEN J

; APPLICANT: ST. GEME III, JOSEPH W

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/038,682

; FILING DATE: 16-MAR-1993

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: BERKSTRESSER, JERRY W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-293

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4937 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-038-682-3

Query Match 1.9%; Score 39.2; DB 1; Length 4937;

Best Local Similarity 53.2%; Pred. No. 0.34;

Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTGAAAAAACTACTCAAGTTGGATGTGAATGCATCCCGCAGGCTATTGAAGA 862

Db 4287 GGTATTAAACGCAAAAGATGCTAAGCTAAATGGTGTATGCATCAGGTGATGATACAGAAGT 4346

QY 863 TGCTGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATCCACGGAAGGATTGTTAA 922

Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGA 4406

QY 923 TATGGATGAAGATGAAGAAACTGAACATGAACAAAA 958

Db 4407 TATCACTGGGGATTAAACACAGTAAATGGGTAAA 4442

RESULT 10

US-08-302-832-3

; Sequence 3, Application US/08302832

; Patent No. 5603938

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-3

Query Match 1.9%; Score 39.2; DB 1; Length 4937;
Best Local Similarity 53.2%; Pred. No. 0.34;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTATTGAAAACTACTCAAGTTGGATGTGAATGCATCCCGGAGGGTATTGAAGA 862
Db 4287 GGTATTAAACGCAAAAGATGCTAAGCTAAATGGTGTGATGCATCAGGTGATAGTACAGAAGT 4346

QY 863 TGCTGAAGAAACAGCAACTCAAACCTTGTTGGTGGGACAGATTCCACGGGAAGGATTGTTTAA 922
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGTTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

QY 923 TATGGATGAAGATGAAGAAACTGAACATGAACAA 958
Db 4407 TATCACTGGGGATTATAAACACAGTAAATGGTTAAA 4442

RESULT 11
US-08-530-198-3
; Sequence 3, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-530-198-3

Query Match 1.9%; Score 39.2; DB 2; Length 4937;
Best Local Similarity 53.2%; Pred. No. 0.34;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTATTGAAAACTACTCAAGTTGGATGTGAATGCATCCCGGAGGGTATTGAAGA 862
Db 4287 GGTATTAAACGCAAAAGATGCTAAGCTAAATGGTGTGATGCATCAGGTGATAGTACAGAAGT 4346

QY 863 TGCTGAAGAAACAGCAACTCAAACCTTGTTGGTGGGACAGATTCCACGGGAAGGATTGTTTAA 922
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGTTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

QY 923 TATGGATGAAGATGAAGAAACTGAACATGAACAA 958
Db 4407 TATCACTGGGGATTATAAACACAGTAAATGGTTAAA 4442

RESULT 12
US-08-469-880-3
; Sequence 3, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:



```
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-3

Query Match      1.9%; Score 39.2; DB 2; Length 4937;
Best Local Similarity 53.2%; Pred. No. 0.34;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY      803 GCTTATTATGAAAAAAGTCTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGA 862
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      4287 GGTATTAAACGCAAAAGATGCTAAGCTAAATGGTGAATGCATCAGGTGATGATACAGAAGT 4346

QY      863 TGCTGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATTCCACGGAAGGATTGTTAA 922
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

QY      923 TATGGATGAAGATGAAGAAACTGAACATGAACAA 958
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      4407 TATCACTGGGGATTAAACACAGTAAATGGGTTAAA 4442
```

```
RESULT 13
US-08-728-470-3
; Sequence 3, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
```

```
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-728-470-3

Query Match      1.9%; Score 39.2; DB 2; Length 4937;
Best Local Similarity 53.2%; Pred. No. 0.34;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY      803 GCTTATTATGAAAAAAGTCTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGA 862
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      4287 GGTATTAAACGCAAAAGATGCTAAGCTAAATGGTGAATGCATCAGGTGATGATACAGAAGT 4346

QY      863 TGCTGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATTCCACGGAAGGATTGTTAA 922
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

QY      923 TATGGATGAAGATGAAGAAACTGAACATGAACAA 958
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      4407 TATCACTGGGGATTAAACACAGTAAATGGGTTAAA 4442

RESULT 14
US-08-617-697-3
; Sequence 3, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

